If. . . : A Theory of Conditionals

Cian Dorr and John Hawthorne

Draft of 1st March 2018

Note for Mind and Language seminar participants: some apologies are in order, for (i) the fact that this is in pretty rough shape; (ii) the fact that it is not so directly about 'semantic frameworks', and (iii) the fact that there is so much of it. We expect to focus on Chapter 1; parts of Chapter 2 are included for context.

Contents

Introduction

There is a long tradition of treating 'if' (or perhaps 'if... then...') as a binary connective, expressing a function from ordered pairs of propositions to propositions. The proposition expressed by a particular 'if' sentence on a given occasion is the result of applying the function contributed by 'if' to two other propositions, the antecedent and the consequent, which are expressed on that occasion by the subordinate clause and the main clause respectively. For example, the proposition expressed by

(1) If Jack is in the park now, Jill is in the park now

on an occasion of utterance is the result of applying the function contributed on that occasion by 'if' to the propositions contributed on that occasion by 'Jack is in the park now' and 'Jill is in the park now'. The subordinate and main clauses needn't be sentences that would serve as standalone declarative utterances. For example, in

(2) If I resigned tomorrow, I would be hired by Google the day after tomorrow

'I resigned tomorrow' would not make sense to assert on its own: nevertheless, it is natural to think of it as contributing a proposition, namely the same one that would be expressed by a standalone utterance of 'I will resign tomorrow'.[1]

[1]We use 'antecedent' and 'consequent' for propositions. Others use these words for linguistic items, either what we call the 'subordinate clause' and 'main clause' of the 'if'-sentence (e.g. 'I resigned tomorrow' and 'I would be hired by Google the day after tomorrow' in the case of (2)), or certain standalone declarative sentences derived by applying certain transformations to these clauses (e.g. 'I will resign tomorrow' and 'I will be hired by Google the day after tomorrow' in the case of (2)). Note that expressions like 'if Jack is in the park now' and 'if I resigned tomorrow', that include the word 'if', are also grammatically clauses—we will call these expressions 'if'-clauses, and reserve the word 'subordinate clause' for the material following 'if'.

This model is prima facie plausible for many 'if'-sentences. But there are some 'if'-sentences for which it is completely hopeless. Consider

(3) If a farmer owns llamas, the farmer is rich

What two propositions would be the inputs to the function contributed by 'if' on this occasion? Certainly 'the farmer is rich' does not contribute the proposition that the one and only farmer is rich, or the proposition that the one and only farmer with such-and-such feature is rich, or a proposition concerning the richness of any particular farmer. If we were determined to force this sentence into the two-proposition model, the best option would be to take the consequent to be the proposition that all the llama-owning farmers are rich, and the antecedent the proposition that there is a llama-owning farmer. But this seems a stretch; and it is far from obvious what linguistic mechanisms would be responsible for associating those two propositions with the subordinate and main clauses of (3).

The dominant strand in the philosophical literature deals with sentences like (3) by silently ignoring them. But there are other cases which are arguably like (3) but which sometimes have been shoehorned into the two-proposition model, such as (4):

(4) If Jack sees Jill next week, he will wave

Here there is a natural candidate to be the antecedent, namely the proposition that Jack will see Jill next week. But what would be the consequent? The proposition that Jack will wave sometime in the future? The proposition that Jack will wave sometime next week? More plausible candidates are the proposition that every time Jack sees Jill next week he will wave at Jill, or the proposition that on at least one occasion next week Jack will see and wave at Jill. But these *ad hoc* reconstructions are not much use to systematic theorising; and it would be a mistake to take it for granted that a systematic theory will conform to the two-proposition model at all.

The two-proposition model is also questionable for conditionals whose main clause contains a modal:

(5) a. If you steal from your employer, you should only steal a little.

b. If he is in the pub, he can't be at home.

c. If he is in the pub, he might be at home.

While the application of the two-proposition model to these sentenecs is not out of the question, there is pressure to think that the consequent is something rather different from the proposition that would be naturally be expressed by a standalone occurrence of the main clause. There is also a tradition in which certain sentences like these are reconstructed in such a way that there is a mismatch between the real and apparent scopes of the modals which on the surface occur in the main clause, so that the content of (5b), for example, would be more perspicuously represented by 'It can't be that (if he is in the pub, he is at home)'. And some (e.g. Kratzer 1986) reject the application of the two-proposition model to these sentences in more radical ways.

One other kind of problem case for the two-proposition model is exemplified by

(6) If Fred had eaten any kind of shellfish, he would have been sick

Some will be comfortable assimilating (6) within the two-proposition model, claiming that its antecedent is the existential proposition that Fred eats some kind of shellfish (during the relevant time period). But we are not sure this is right: thinking about the difference between the natural reading of (6) and the natural reading of (7),

(7) If Fred had eaten any kind of shellfish, he would have eaten lobster

one feels that the subordinate clause in (7) is much more apt than the one in (6) for a mundane existential treatment. (The role of 'any kind of shellfish' in (6) strikes us as more reminiscent of role of 'llamas' in (3) above.) Moreover, the reasons for being cautious in the case of (6) arguably carry over to (8), whose contrast with (9) is similar to that between (6) and (7):

(8) If Fred had eaten either lobster, crab, or prawns, he would have been sick.

(9) If Fred had eaten either lobster, crab, or prawns, he would have eaten lobster.

An adequate theory of conditionals must account for all these sentences. Nevertheless, in the first four chapters of this book, we will focus entirely on conditionals like (1) and (2), where it is espeically plausible that the meaning somehow involves two propositions, and where there is no relevant uncertainty about which propositions they are. We will defend a view on

which these sentences express propositions that are evaluable for truth and falsity, and we will attempt to say something systematic about the kind of relation that needs to hold between the antecedent and the consequent for the proposition expressed by the conditional to be true. In developing this theory we will not need to make any particular assumptions about the syntactic structure of conditionals or about the semantic contribution of the word 'if'. For example, it is compatible with our view that 'if' doesn't have a semantic value at all, or has a trivial semantic value, so that the semantic value of the 'if'-clause 'if Jack is in the park now' is the same as the semantic value of 'Jack is in the park now' (just as, e.g., 'snow is white' and 'that snow is white' arguably have the same semantic value in 'I believe that snow is white'). On this picture, the determination of the semantic value of the whole conditional sentence will be governed entirely by compositional rules and/or the semantic values of unpronounced constituents.[2] Still less are we making any assumption about the semantic value of 'then' in 'If *P* then *Q*', or committing ourselves to the unlikely hypothesis that 'if' and 'then' comprise a single scattered semantic unit. All that matters for our positive view is that the contributions of the subordinate and main clauses in (1) and (2) are propositions (or things that in some natural way determine propositions), and the very propositions that the philosophical tradition takes them to be.

Having narrowed our field in this way, we are in a position to state the theory we will be defending in a schematic way, as follows:

CLOSEST A conditional with antecedent p and consequent q is true iff either there is no accessible world at which p is true, or the closest accessible world at which p is true is a world at which q is true.

The schema is anything but new: in fact it is one of the oldest in the book, having been around since the seminal works of R. C. Stalnaker 1968 and Lewis (1973). Our contribution will be to offer certain ways of fleshing out the key terms of art: 'accessible', 'closest', and 'world'. Obviously, different ways of cashing out this ideology will yield wildly different theories. Suppose for example that we interpret 'accessible' in such a way that at any given world, that world is the only accessible world. Then the schema will yield the same truth conditions as the material conditional: when the antecedent is false, it is false in all accessible worlds, and so the conditional

[2] Examples like 'If he came to the party and if he behaved himself, he was invited back' are especially problematic for the idea that the semantic value of 'if' is a relation between propositions.

is vacuously true; when the antecedent is true, the actualised world must be the closest accessible world where the antecedent is true (since it is the only accessible world), and so in that case the conditional is true just in case the consequent is true. Our preferred take on the relevant ideology, which works very differently from this, will emerge over the course of the next four chapters. chapter 1 will argue that accessibility is highly context-sensitive (to the extent that it is rare for two utterances of conditionals to invoke the same accessibility relation), and often quite demanding, although rarely so demanding that it would make the conditional collapse into a material conditional. chapter 2 will motivate the invocation of a closeness relation for which, as CLOSEST implicitly assumes, there is always guaranteed to be a unique closest world within the set of the accessible worlds where any p is true (so long as it is not empty), and argue that closeness should not be thought of as having anything at all to do with similarity. Chapter 3 will formulate and defend certain generalisations about the objective chances of closeness-theoretic propositions and the credences they deserve, with a view to explaining why the conditional chance or credence of the consequent of a conditional on its antecedent is often a good guide to the chance or rational credence of the proposition expressed by the conditional. Chapter 4 will consider some cases which are problematic for CLOSEST given familiar accounts of possible worlds, and consider to what extent they can be solved by countenancing impossible worlds. Chapter 5 will return to the cases we have just been setting aside, where the application of the two-proposition model is contested, and argue that they can be handled in a way that does not disrupt the idea that CLOSEST belongs at the core of the semantic theory of conditionals. Finally, the Conclusion will address the worry that our account makes our ordinary practice of reasoning with conditionals hostage to strange, spookly committments in metaphysics and epistemology, and holds out the hope that the committments in question can be made palatable by appealing to independently-motivated ideas about vagueness.

Chapter 1

Accessibility

1.1 The context-sensitivity of conditionals

Even holding fixed which propositions are the antecedent and consequent, there are different propositions that could be expressed by uttering a conditional on different occasions. For one thing, it is well known that the difference between 'indicative morphology' and 'counterfactual morphology' is semantically significant, as exemplified by Ernest Adams's famous minimal pair:

(1) a. If Oswald didn't kill Kennedy, someone else did

b. If Oswald hadn't killed Kennedy, someone else would have

This difference seems to be driven by something about the pattern of tense markings that appear on the surface within the main and subordinate clauses; however, the way in which these morphological differences actually contribute to meaning is quite controversial. For the moment, we will take a capacity to sort "indicative" from "counterfactual" conditionals for granted, and we will try only to use examples whose classification is uncontroversial.

For another thing, the proposition expressed by a conditional can vary even when the wording remains exactly the same. In the case of counterfactuals (i.e. sentences syntactically like (1b)), this variability can be illustrated by an example of Quine's (described by Lewis 1973):

(2) a. If Caesar had fought in Korea, he would have used nuclear weapons

b. If Caesar had fought in Korea, he would have used catapults

Given appropriate conversational background, it seems that one could speak the truth by uttering each of these sentences, though in no normal background could one speak the truth by uttering their conjunction. Similarly,

each of (3a) and (3b) (Jackson 1977) seems like something one could use to assert a truth, although their conjunction seems absurd:

(3) a. If I had jumped out the window right now, I would have been killed

b. If I had jumped out the window right now, I would have done so only because there was a soft landing laid out for me

In the case of indicatives (i.e. sentences syntactically like (1a)), the case for context-sensitivity is more controversial, but still strong. The main observations here are due to Gibbard (1981), although the following case is due to Bennett. The setting is one where there are two channels by which water can flow from an upper reservoir to a lower reservoir; each is closed off by a gate. Speaker A discovers that the east gate is closed and asserts

(4) If the water got to the bottom, it got there via the west channel

Speaker B discovers that the west gate is closed and asserts

(5) If the water got to the bottom, it got there via the east channel

Intuitively, both speeches seem true. But the conjunction of (4) and (5) is extremely odd. It is thus very natural to explain the acceptability of (4) and (5) by appeal to context-dependence.

We propose that all of the differences we have just been talking about are to be explained by differences in what it takes for a world to be "accessible" in the sense relevant to the context. For example, the accessibility relation operative in the context where (2a) seems true requires match with respect to what kinds of military technology is in use at a given historical period, whereas the one operative in the context where (2b) seems true does not require this, but does require match with respect to what kinds of military technology were actually available to particular people.[1] Similarly, in the context where (3a) seems true, accessibility requires match with respect to the laws of nature and the course of history up until very shortly before the

[1] Note that if, unbeknownst to us, there are no such things as nuclear weapons and catapults are still in wide use even now, the true-seeming utterances of (2a) are in fact false; similarly, if unbeknownst to us, Martians actually provided Caesar with nuclear weapons which he only didn't use because he had no need to resort to them, the true-seeming utterances of (2b) are in fact false.

time of utterance, whereas in the context where (3b) seems true, accessibility requires much less than this.[2] In the contexts where (4) and (5) seem true, meanwhile, accessibility involves something like compatibility with the knowledge of the speaker at the time of utterance, and the contextual variability is primarily due to variation in who is speaking and when.

Even holding fixed who is speaking and when, we think there is plenty of room for further contextual variation in the interpretation of the accessibility parameter for indicative conditionals (like (4) and (5)). First, in some cases the knowledge state of other relevant individuals or groups may be relevant, rather than just the knowledge of the speaker. Secondly, the epistemic relation that fixes accessibility may not always be knowledge—in some settings it may be something more demanding (e.g. being known for sure), while in others it may be laxer. Thirdly, when questions are salient (or 'under discussion') in a conversation, they may serve to constrain accessibility, so that the worlds accessible from any given world are required to match that world with respect to the answers to those questions, in addition to being epistemically live. (The phenomenon of constraint by questions under discussion also arises for counterfactuals.) We will return later to all of these potential dimensions of contextual variation, and put them to work in explaining various data points.

The structure of CLOSEST does not *require* attributing the differences we have been considering to differences in the accessibility parameter. One could instead attribute some or all of them to differences in the operative closeness relation, in which case one could, if one wished, hold that the accessibility parameter is always wide open (i.e. that every world always counts as accessible from every other). Proponents of this approach could even mimic our specific suggestions about what plays the accessibility role in particular cases by saying that while the operative accessibility relations do not vary, the operative closeness relations vary in such a way that each world we would count as accessible counts in context as closer than each world we would count as inaccessible. Much of our picture could survive this change in the division of labour between accessibility and closeness; however, the picture of the accessibility parameter as the primary driver of context-sensitivity has some advantages which will emerge in the present chapter.

[2]The relevant notion of "match" here may not require *perfect* match in all respects, no matter how microscopic: see Dorr 2016.

1.2 The difference between indicatives and counterfactuals

The grammatical difference between indicative and counterfactual conditionals clearly has some systematic semantic significance: given what we claimed the previous section, this must consist in some systematic difference in the values that are allowed for the accessibility parameter. Loosely following von Fintel (1998), we suggest that this difference takes the form of general constraint on indicatives which does not apply to counterfactuals: namely that for an indicative conditional, *accessibility must entail epistemic possibility*. Roughly, to say that a world is epistemically possible is to say that it is a live candidate for being how things actually are, from the perspective of some contextually relevant individual or group, typically the speaker or a group containing the speaker (where liveness for a group is understood as entailing liveness for each member of the group).[3]

This explains why we are forced to use counterfactual morphology to get across the plausible thoughts we naturally would try to get across by uttering (2a) or (2b). In the contexts they naturally evoke, these sentences are non-vacuously true—that is, (2a) evokes a context where 'If Caesar had fought in Korea, he would never have used nukes' is false, and (2b) evokes a context where 'If Caesar had fought in Korea, he would never have used catapults' is false.[4] It is very hard to get oneself into a mindset where the proposition that Caesar fought in Korea is regarded as epistemically "live". (And it is harder still to get oneself into a mindset where there are "live" worlds in which Caesar was present in Korea but where one is still willing to assert something that entails that Caesar didn't use nukes in Korea or something

[3]von Fintel (1998) instead proposes that for indicatives, the accessible worlds must be a subset of the "context set" in the sense of R. C. Stalnaker 1978—the set of worlds consistent with what everyone in the conversation "commonly presupposes". This weaker demand can't do as much work as the notion of liveness we are working with. For example, the liveness constraint readily explains why it's appropriate to say 'If John didn't bring his umbrella today he got wet' when you can see it's raining outside but your audience has no idea it's raining: you know that there are no live worlds where John didn't bring his umbrella and didn't get wet, so you're in a position to know that no worlds like this are accessible (in the relevant sense). By contrast, since there are some worlds like this in the common ground, something would need to be added to the "context set" constraint to explain why the speech is good. If one wants to understand the relevant notion of epistemic possibility in terms of common presupposition, a more promising idea is to follow Mandelkern in taking the relevant set to be the *prospective* context set—the set of worlds that would remain in the contextset if the sentence being uttered were accepted.

[4]This is explained in part by the presupposition of nonvacuity, to be discussed in section 1.3 below.

that entails that Caesar didn't use catapults in Korea.) Similarly, conditionals beginning with 'If I didn't exist right now...' are generally much more intelligible than those beginning with 'If I don't exist right now...': making sense of the latter requires entering into a most unusual sense of openness to nihilistic metaphysics.

Should we posit a correlative constraint for counterfactuals, according to which the value of the accessibility parameter for a counterfactual has to be a property of worlds that does *not* entail epistemic liveness? The generalisation that counterfactuals are in fact interpreted using such accessibility relations looks quite plausible. Often, the antecedents of counterfactuals are clearly not regarded as epistemically live, in which case the only way for them to be nonvacuously true is for some worlds that are not epistemically live to be accessible. And even when the speaker regards the antecedent of a counterfactual as epistemically live, the presence of non-epistemically live worlds among the accessible ones can be crucial, as in this famous example:

(6) If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show. (Anderson 1951)

As von Fintel (1998) notes, the speaker obviously knows that Jones is showing just exactly those symptoms which he does in fact show, so the conditional will be *obviously*, and uninterestingly, true if we interpret it using a notion of accessibility that entails compatibility with the speaker's knowledge. However, it's not clear that we need to give the generalisation that accessibility for counterfactuals does not entail liveness the same rule-like semantic status as the generalisation that accessibility for indicatives does entail liveness. If we treat one of the generalisations as a rule, we could perhaps recover the other one using some general principle enjoining cooperative speakers to choose the more constrained of two possible forms when possible.[5] We will not take a stand on whether this should be done.

The idea that accessibility for indicative conditionals entails epistemic liveness fits naturally with an account of epistemic modals: 'It must be that *P*'; 'It might be that *P*'; 'It is possible that *P*', etc. In orthodox fashion, we take these to be context-sensitive, in a way that can be represented by an

[5] If we implemented one of the generalisations as a *presupposition* (that accessibility entails or fails to entail epistemic liveness), we could derive the other via the principle *Maximise Presupposition* (Heim 1991), according to which when one of two sentences with the same "assertive content" has a stronger presupposition, speakers are expected to utter that one when its presupposition is in fact common ground in the conversation.

accessibility parameter: 'It must be that *P*' is true iff *P* is true at all accessible worlds, and 'It might be that *P*' and 'It is possible that *P*' are true iff *P* is true in some accessible world. The "epistemic" character of these modals consists in the fact that accessibility always has to entail epistemic liveness (from the standpoint of a contextually relevant individual or group). Moreover, when one uses both epistemic modals and indicative conditionals, the default is for both to be interpreted uniformly (using the same notion of accessibility). This means that the argument-schemas

MUST-IF It must be that *Q*. Therefore, if *P*, *Q*.

and

MIGHT-PRESERVATION It might be that *P*. If *P*, *Q*. Therefore, it might be that *Q*.

are both valid, in the sense that on any uniform interpretation, the propositions expressed by the premises entail the proposition expressed by the conclusion. Arguments of these forms certainly *seem* valid, so getting them to come out valid is a significant advantage of the approach that appeals to variation in the accessibility parameter to explain the relevant context-sensitivity over the competing approach mentioned in the previous section that keeps accessibility constant and appeals to variation in the closeness relation. Moreover, as we will see in section 1.4, the validity of MUST-IF helps to explain certain facts which have often been used to argue for competing views such as the theory that indicative conditionals are material conditionals.

One might worry that the proposed constraint on accessibility for indicative conditionals is too demanding. Since an utterer of (1a) ('If Oswald didn't kill Kennedy, someone else did') is clearly not presenting it as merely vacuously true (for reasons to be explored in section 1.3), the operative notion of accessibility must be one on which they are assuming that there are accessible worlds where Oswald didn't kill Kennedy; given the proposed constraint, this means the speaker must be regarding the possibility that Oswald didn't kill Kennedy as epistemically live. This seems strange: after all, most of us know perfectly well that Oswald did kill Kennedy and yet are happy to utter (1a), and not simply in the spirit in which we might utter 'If Oswald didn't kill Kennedy I'm a monkey's uncle'. There are a few different routes one might follow in responding to this objection. On a radical approach, the proposition expressed by utterances of (1a) by us

knowledgeable folk is in fact vacuously true, but when we utter it, we take for granted the false proposition that it isn't. (This could be a kind of pretended epistemic modesty, or else a briefly held false belief that we don't really know who shot Kennedy.) Similarly, the radical view will say that despite the appeal of 'I might be dreaming' is natural, utterances of it will express a false proposition that is treated as acceptable either because we are pretending that it is true, or because we falsely believe that it is true (since at the time of utterance we are taken in by the thought that we don't know whether we are dreaming). On a less error-theoretic view, (1a) would be nonvacuously true in the context that would be evoked if one of us were to assert it, and 'It's possible that Oswald didn't killed Kennedy' is also true by the standards of this context. On this approach, what counts as "live" in the sense relevant to the constraint is itself a contextually variable matter, and speeches like (1a) push us towards an especially permissive resolution of this context sensitivity. One might want to link this contextual variation to context-sensitivity in the verb 'know', saying that 'We don't know whether Oswald killed Kennedy' expresses a truth in the relevant contexts (although it expresses a falsehood in many others). Alternatively, one might conceive of liveness as setting a contextually flexible epistemic standard that does not always march in lock-step with the one associated with 'know': for example, liveness could in some cases be a matter of consistency with some more attenuated body of knowledge, or liveness might be a matter of consistency with a body of propositions of which one has an especially secure kind of knowledge.

Epistemic modals need not always be anchored to the facts about what is live at the time of speech. In a sentence like 'It was possible that it had rained, since the ground seemed a bit damp', the operative epistemic perspective is in the past: what we are saying, roughly, is that what the relevant people knew at the relevant past time was consistent with the proposition that it had rained earlier than that time. There is no suggestion that the speaker is ignorant or in any way open at the time of speech to the possibility that it had rained: what matters is just the epistemic standpoint of the salient person or group (or more abstract 'perspective') as of the target time.[6] 'It was possible that it was raining' and 'It was possible that it was going to rain' are similar. 'Might have' and 'could have' claims also have a reading

[6] In some cases one needs to work with such notions as 'the evidence that was available at the time' even when there is no relevant person around to gather it: ** Add example about the early universe **

that works like this: 'It might have been raining' can mean 'It was possible that it was raining', although it can also mean 'It is possible that it has been raining'.[7]

Given the intimate relation between epistemic modals and indicative conditionals, we would thus expect that in some cases, the accessibility parameter of an indicative conditional is tied to an earlier epistemic perspective. This does indeed seem to be going on in examples like the following:

(7) a. If Oswald hadn't killed Kennedy, someone else had
b. If Oswald hadn't killed Kennedy, someone else would
c. If Oswald didn't kill Kennedy, someone else would
d. If Oswald wasn't killing Kennedy, someone else was

Just like 'It was raining' and 'Einstein was going to win a Nobel Prize', these are sentences that could not be felicitously asserted out of the blue: some particular past time has to be salient, and the proposition asserted is in some way about that time. (The required salience could be achieved either by earlier discourse or by nonlinguistic clues.) In the case of (7a)–(7d), the most natural reading is one on which one of the roles of this salient past time is that of providing the relevant epistemic perspective. None of these sentences commits the speaker to accepting 'It *is* possible that Oswald didn't kill Kennedy', or to regarding the proposition that Oswald didn't kill Kenedy as a live possibility: but they do seem to commit us to a claim about what *was* epistemically possible at the time in question. Note that one natural use for sentences like (7a)–(7d) is in indirect speech reports of present-tense speeches made at the relevant time, namely

(8) a. If Oswald hasn't killed Kennedy, someone else has
b. If Oswald hasn't killed Kennedy, someone else will
c. If Oswald doesn't kill Kennedy, someone else will
d. If Oswald isn't killing Kennedy, someone else is

Given the account we have outlined, this is to be expected, since the very propositions that would be asserted by (8a)–(8d) can be expressed at later times by (7a)–(7d).

[7] In some other natural languages where the analogues of 'might' are normal tensed verbs, this ambiguity is lexically resolved. In English 'have to' works like this: we distinguish 'It has to have been raining' from 'It had to be raining'.]

Although there are some syntactic similarities between the likes of (7a)–(7d) and (1b) ('If Oswald hadn't killed Kennedy, someone else would have'), there is a significant semantic gulf between them. On its natural interpretation, (1b) does not commit us to its being a live possibility that Oswald didn't kill Kennedy, either from our own perspective or from any other perspective—intuitively, its meaning feels altogether more "worldly" by comparison with the "perspectival" feel of (7a)–(7d). This is what we have been getting at in our use of the labels 'indicative' and 'counterfactual', roughly in line with the philosophical tradition. Note however that examples like (7a)–(7d) are rather different from the paradigmatic examples of indicative conditional sentences, and would provide counterexamples to many generalisations that philosophers have been wont to make about indicative conditionals (e.g. generalisations about how the probability or "assertability" of an indicative conditional relates to the corresponding conditional probability), or about the superficial linguistic form distinctive of indicatives and/or counterfactuals (for example, (7b) shows that a 'would' in the main clause does not suffice for counterfactuality).

In fact, once we recognise the category of past-perspective indicative conditionals, we can see that there is no failsafe way to distinguish counterfactuals from indicatives on the basis of superficial syntax, since many sentences admit of both kinds of interpretation, including the paradigm counterfactual (1b):

(1b) If Oswald hadn't killed Kennedy, someone else would have

Although the dominant reading of this sentence is certainly counterfactual, it also has a past-perspective indicative meaning, which comes to the fore when we imagine using (1b) in reporting a past utterance of the following somewhat unusual but perfectly intelligible sentence:

(9) If Oswald hasn't killed Kennedy, someone else will have

And once this epistemic interpretation has been noticed, one can imagine contexts in which it is the intended interpretation even outside indirect speech reports.[8]

There are, however, some conditionals for which a counterfactual meaning is grammatically required. The most prominent examples are ones in which the main verb in the 'if'-clause takes a subjunctive form:

[8]Khoo (2015: p. 15) also argues that (1b) has an "epistemic" reading.

(10) a. If Gore were president, he would deal with this problem

b. If I were to resign tomorrow, I would be hired by Google the day after tomorrow.

Putting certain present-tense verbs in the main clause also rules out indicative readings:

(11) a. If I resigned tomorrow, I reckon I would be hired by Google the day after tomorrow.

b. If I got a tattoo, it is unlikely that anyone would notice.

It is plausible that the unavailability of indicative readings can be traced to the same source as the unacceptability of sentences like 'It was possible that I am happy'.

There are also cases where a counterfactual reading seems to be required on semantic grounds:

(12) a. If I had married someone other than the person I did marry, I would not be happy.

b. If giraffes were any taller than they actually are, they wouldn't be able to pump blood up to their brains

In these cases, it is plausible that what forces the counterfactual reading is the fact that the antecedent is so obviously non-live.[9]

Iatridou (2000) argues compellingly that what is distinctive of counterfactual uses of conditionals (in a range of different languages) is what she calls 'fake past tense': an extra layer of past tense morphology that does not carry the usual significance of temporal pastness. Our observation that canonical counterfactuals like (1b) can also be interpreted as past-perspective indicatives supports Iatridou's claim. When the sentences are used as indicatives, their past tense morphology manifestly *does* have its usual temporal meaning, with the result that these uses are felicitous only when there is a salient past time for them to refer to; by contrast, when understood as counterfactuals they are perfectly fine out of the blue.

As Iatridou notes, apparently non-temporal uses of past-tense morphology also crop up in certain other environments, as in the following examples.

[9]When (12a) and (12b) are embedded they need not always be counterfactuals: for example, 'He said that if I had married someone other than the person I did marry, I would not be happy' could be a felicitous report of a past utterance of the 'If he has married someone other than Jessica, he will not be happy'.

(13) a. I wish they were here right now.
b. Oh, that they had been here right now!
c. They might have been here right now.
d. They couldn't have been here right now.

'Might have been' and 'could have been' also have epistemic uses where the past tense is non-fake—as noted earlier, 'It might/could have been raining' can mean 'It is possible that it has been raining' and 'It was possible (then) that it was raining (then)'. But the use of 'right now' in (13c) and (13d) seems to preclude these readings, presumably for reasons similar to those that make for the badness of 'It is possible that they have been here right now' and 'It was possible that they are here right now'.

In many paradigmatically counterfactual conditionals, the 'if'-clause involves pluperfect morphology, which can be thought of as involving two "layers" of past tense. In an ordinary use of the pluperfect both layers play a temporal role: 'I had eaten breakfast' takes us to a past "reference time" and then places an eating event earlier than that time. However, as Iatridou notices, there is no similar sense of double pastness in the natural counterfactual use of 'If I had eaten breakfast this morning, I would have skipped lunch'. Her view of these cases is that at least one layer of past tense is fake past, while at most one is the usual temporal past. (In 'If it had been raining right now, the ground would have been wet', both layers seem to be fake.)

Why should the past tense be ambiguous in this way? An intriguing but elusive idea of Iatridou's is that the past tense has a more skeletal core meaning of "distance", which can be cashed out either temporally or modally. In our framework, the relevant thought in the modal case would presumably be one to the effect that non-live possibilities are "distant", so that invoking a notion of accessibility that extends to the non-live requires reaching out to the distant. Obviously this picture raises many questions—for example, why do actual *future* events not count as "distant" and thus apt to be described by verbs in the past tense? But we won't try to address such questions here: as with many other facts about linguistic structure, recognising that the phenomenon of fake past tense exists does not require having an explanation of why language is so configured.

We have posited a interpretative link between indicative conditionals and epistemic modals: barring context-shift, they invoke the same accessibility relation. Given that fake past tense can occur with modals, as in (13c) and (13d), it is plausible that these modals stand in an analogous relation to

counterfactual conditionals. If so, the non-epistemic reading of 'It couldn't have been that P and Q' will (holding context fixed) entail 'If it had been that P, it would have been that not Q'. Also, 'It could have been that P' and 'If it had been that P, it would have been be that Q' will jointly entail 'It could have been that Q'.These claims seem no less plausible than the corresponding claims about indicatives.

There is a competing picture of the role of the past-tense morphology in counterfactual conditionals which rejects the idea of 'fake past', and instead regards the relevant uses of the past tense as introducing reference to genuine past times (just as it uncontroversially does in past-perspective indicative conditionals like (7a)–(7d)). As developed by Khoo (2015), following Condoravdi (2002), the idea is that just we can talk about a possibility being epistemically live or not at a given time, we can talk about a possibility being "metaphysically live" or not at a given time, where the metaphysically live possibilities at a time are those which share their history up to that time. The picture is that when we utter a counterfactual, there is a particular time earlier than the time of utterance such that the accessible worlds are all and only those that are metaphysically live at that time, and at least one of the past-tense morphemes in the conditional refers to that past time. Khoo offers this as an explanation for why the grammatical differences between standard indicatives and counterfactuals matters to interpretation: the picture is that (1a) has to have an epistemic meaning since it its conditional has to be interpreted with respect to the present time, but because its antecedent is about the past it is guaranteed to be true either at all or none of the worlds that are metaphysically live at the present time: if we want a conditional whose accessibility parameter applies to all the worlds metaphysically live at some time where some but not all of these worlds are worlds where Oswald didn't kill Kennedy, the time in question will need to be in the past, so we will have to add some past tense morphology to let us refer to it.

There are several problems with this view. First, it will have trouble explaining why standard counterfactuals can be acceptable even when no particular past time has been raised to salience as a target to be referred to by the relevant morphemes—by contrast, 'It was raining' clearly requires such salience to be in place, and so do past-perspective indicatives like (7a)–(7d). Second, the view fails to explain why an "epistemic" reading is the only one available for (7c) ('If Oswald didn't kill Kennedy someone else would'), given that its reference time clearly has to be earlier than the killing.[10] Third,

[10] Also, in the past-perspective indicative reading of (1b), as in (7a) and (7b), the "reference

the view struggles with counterfactuals about the future like 'If I resigned next week I would be hired by Google the week after'. Even granting that the set of accessible worlds consists in all those whose history matches that of the actual world up to some given time, there is little reason to think that the time in question is in the past: it is much more plausible that the time of divergence is identical to or after the time of utterance, given that in evaluating such counterfactuals, we freely draw on known truths about the history of the actual world right up to the time of utterance. Fourth, it is just not plausible the notion of accessibility relevant to the evaluation of counterfactuals always requires matching any part of the history of the actual world: consider 'If gravity had obeyed an inverse cube law, stars would have been unstable', 'If there had always been infinitely many stars, then there would always have been infinitely many planets', etc. Fifth and finally, it is very hard to see what reference to a past time could be going on in 'I wish he were here right now'; but once we admit that the past tense is fake here, what reason is there to think it is genuinely temporal in 'If he were here right now, then he would be happy' (especially considering that the word 'were' can do double duty in 'If, as I wish, he were here right now, he would be happy').

1.3 The presupposition of non-vacuity

An obvious worry about the proposal that the interpretation of conditionals typically involves a non-trivial accessibility relation is that, when combined with the view that a conditional is true when its antecedent is true in no accessible world, it makes it easy for conditionals whose consequents are contradictory or otherwise bizarre, or whose consequents contradict their antecedents, to be (vacuously) true. This may seem problematic, given that such conditionals are almost always unacceptable.

We propose an explanation in terms of *semantic presupposition*: a conditional interpreted with a particular accessibility parameter has a presupposition that its antecedent is true in some accessible world. 'If P, Q' thus presupposes something equivalent to what 'Not (If P, Q and not Q)' semantically expresses. Also, assuming the accessibility parameter for a modal 'might' or 'could have' is interpreted uniformly with the conditional, 'If P, Q' presupposes what 'It might be that P' or 'It could have been that P' expresses. Conditionals whose consequents contradict their antecedents are thus guaranteed to either express or presuppose something false: this explains why they are bad to assert.

While we expect speakers to strive to avoid uttering sentences with false presuppositions just as they strive to avoid uttering sentences which express false propositions, presuppositions are distinctive in that they are "in the background" rather than being "at issue". In uttering a sentence with p as a semantic presupposition, one signals that p is not just true, but something appropriately *taken for granted*. Often, what makes this appropriate is that p is something ones interlocutors are *already* taking for granted; otherwise, if they are trusting (or careless), they will often begin taking p for granted upon receiving the signal (the process of "global presupposition accommodation"). This makes presupposition a particularly helpful tool for driving context-shift (non-uniformity among different occurrences of context-sensitive expressions). For example, if we have been working with a relatively narrow notion of accessibility and I come out with something like 'If Oswald didn't kill Kennedy, someone else did', listeners are not so likely to conclude that I regard Oswald's not having killed Kennedy as a live possibility in the demanding sense (since even if I held this surprising view, it would be foolish for me to assume the listeners' willingness to take it for granted); more likely, they will instead conclude that I must have intended an interpretation using a new, broader accessibility parameter on which it's uncontroversial that there are accessible worlds where Oswald didn't kill Kennedy. By contrast, if I come right out and say 'Oswald might not have killed Kennedy', I am more likely to be interpreted as making the controversial claim that Oswald didn't kill Kennedy in some world accessible in the old sense, rather than saying something boring and obvious.

The notion of presupposition is often modelled using a trivalent framework where having a false presupposition is equated with being neither true nor false (e.g. Heim and Kratzer 1998). In this framework, positing the presupposition of nonvacuity would require some revisions to the statement of CLOSEST, and a rethinking of a fair amount of what we will be saying about the logic of conditionals. But we will not be using the trivalent system, which raises many foundational issues we would prefer not to have to deal with. Rather, we will treat presupposing and expressing as logically independent relations between sentences and propositions (taken to be

time" must be after the (actual-world) killing of Kennedy. This makes it implausible that we get the counterfactual reading of (1b) just by using a metaphysical rather than epistemic accessibility relation, since the worlds metaphysically live at these post-killing times are all worlds where Oswald did kill Kennedy.

always true or false), so that there is a fourfold classification of sentences with respect to a given interpretation: true with only true presuppositions, true with false presuppositions, false with only true presuppositions, and false with false presuppositions. (Note that this leaves it open whether the presupposing relation can somehow be explained in terms of the expressing relation together with general pragmatic principles, or needs to be taken as an additional component of conventional meaning.) Although most of what we say could survive being transplanted into other ways of thinking of presupposition, we will leave this task to proponents of those frameworks.

In addition to explaining why conditionals whose consequents contradict their antecedents are bad, the presupposition of nonvacuity helps explain why certain inferences which come out valid given CLOSEST in fact seem problematic. One example is the inference from 'It must be that not-*P*' (or 'It can't be that *P*') to the indicative 'If *P*, *Q*', for arbitrary *Q*. Similarly in the counterfactual case, 'It couldn't have been that *P*' will entail 'If *P* it would be that *Q*' for any *Q*. But arguing in this ways seems intuitively bizarre. We can explain this by saying that if accessibility relation for the conditional is interpreted as the same one that matters for the modal, the premise entails that the presupposition of the conditional is violated; so the discourse puts pressure on us to invoke two different accessibility relations, in which case the inference will be simply invalid.

The presupposition can also be motivated in a way that does not depend on any assumptions about the truth values of conditionals whose antecedents are true in no accessible world, by applying standard tests for presupposition. One such test looks for inferences which a sentence gives rise to not only when asserted but when asked as a polar (yes-no) question, and when embedded under various negation-like operators such as 'I doubt that'. The inferences from the indicative 'If *P*, *Q*' to 'It might/could be that *P*'/'it's possible that *P*'/'there's some chance that *P*', and from the counterfactual 'If it were that *P* it would be that *Q*' to 'It could have been that *P*', pass these tests quite well. Consider the oddity of 'I doubt that he's having fun if he's in the pub, and moreover there's no chance that he's in the pub', or the naturalness of responding to 'Would you have agreed if I had bought you a yacht?' with 'I didn't know you could have done *that*'. Another useful test (von Fintel 2004) is the 'Hey wait a minute' test: the inference from *P* to *Q* is presuppositional when it's natural to object to *P* by saying something like 'Hey wait a minute, who said that *Q*?', indicating a refusal to go along with the suggestion that *Q* should be *taken for granted*. Again, the inferences in

question seem to pass the test: consider 'If the Pope comes to my party I'll be delighted. —Wait, is there really a chance that the Pope is going to come?'.

One could instead explain the badness of 'If *P*, *Q* and not *Q*' and the inference from 'It can't be that *P*' to 'If *P*, *Q*' by strengthening the analysis to make a conditional false when its antecedent is true in no accessible worlds. On this view, 'If *P*, *Q* and not *Q*' is in fact contradictory, and the inference-form is invalid. But proponents of the stronger truth-condition still have prima facie reason to accept the presupposition of non-vacuity, stemming from the fact that the inferences survive embedding and the 'Hey wait a minute' test. So making a positive argument for the stronger truth condition based on its capacity to explain these facts will require somehow undermining this prima facie case. The chief disadvantage of the strong truth condition, meanwhile, is that it yields a much less attractive logic than a view on which vacuity makes for truth. For example, *Identity* ('If *P*, *P*') and MUST-IF ('Must *Q*, so if *P*, *Q*') are no longer valid. While the explanatory work that we have done by appealing to these argument-schemas could probably be done by other means (e.g. appealing to statuses like Strawson-validity), we expect the resulting explanations would be more complex and weaker. Overall, the strong truth condition for conditionals is subject to many of the same objections as the view that 'Every *F* is *G*' is false when 'Nothing is an *F*' is true. Once one has admitted the status 'true with a false presupposition', the advantages that these views might seem to bring can be gained at less cost.

[. . .]

As we noted at the beginning of this section, there are some cases where it seems acceptable to utter a conditional (with a particular setting of the accessibility parameter) even though the proposition that the antecedent is true in some accessible world is not known, or even known to be false. 'Monkey's uncle' conditionals are one example; indicatives with absurd consequents uttered as a prelude to an argument by *modus tollens* may be another.[11] This would be problematic for the thesis that conditionals presuppose non-vacuity if we thought that uttering a sentence semantically presupposing a certain proposition was robustly associated with commitment on the speaker's part (in the same way as uttering a sentence semantically *expressing* a proposition). It is often said that presuppositions, or at least certain kinds of presuppositions, are robust in this way, on the basis of the

[11] This is not inevitable: we might instead invoke involve a wide domain of accessible worlds that includes impossible worlds at the absurd consequents are true (see chapter 4).

oddity of speeches like 'The king of France is bald and there is no king of France' or 'He stopped smoking but I don't mean to suggest that he used to smoke'. But the presuppositions in these cases are also entailed by the proposition expressed, so the oddity of these sentences is no surprise—they are contradictory. Once we turn to presuppositions that are *not* also entailments, we see a spectrum of robustness. In some cases, the inferences are quite robust, and attempts at cancellation are pretty befuddling to ordinary speakers: consider 'Every burglar who entered the White House was sent by the FBI, because there weren't any burglars in the White House'. In other cases, the tendency for speakers to assume the truth of the standardly-predicted presupposition seems to disappear altogether in certain contexts, e.g. when certain questions are made salient—consider 'He doesn't *know* it's raining! Remember that he's relying on those notoriously unreliable instruments'. Perhaps the lesson to draw from this variety is that the category of "presupposition" is something of a grab bag within which important distinctions need to be drawn. In any case, the existence of this spectrum means that what we seem to find with conditionals—namely, a generally quite robust association with a few circumscribed exceptions, some conventionally marked—is not especially problematic for, though also not explained by, the presuppositionality hypothesis.

Asserting a conditional doesn't just tend to convey that the antecedent is true in some accessible worlds: it also tends to convey that the *negation* of the antecedent is true in some accessible worlds. Consider: 'If the trains aren't running we can just take a taxi. —Hey wait a minute, I didn't know there was a chance that the trains weren't running!' Should this kind of inference also be explained by positing a presupposition (which, in conjunction with the presupposition of nonvacuity, would amount to a presupposition of 'antecedent diversity', to the effect that the accessible worlds are not all alike with respect to the truth value of the antecedent)? We are not sure. The generalisation that hearers infer 'Might not P' from 'If P, Q' does seem to have many more exceptions that the generalisation that they infer 'Might P': for example, there is no temptation to draw such an inference when 'If P, Q' is uttered immediately after P as a prelude or invitation to a *Modus Ponens* inference (to Q). ('...So the butler did it; but if the butler did it, I'm innocent and deserve to be freed'). A promising alternative way of explaining this inference is to invoke conversational implicature ("competition effects") instead of presupposition, the idea being that if speakers are in a position to assert 'Must Q' we expect them to do so rather than asserting the weaker and

more complex 'If P, Q'. As noted above, it is hard for this kind of reasoning to get beyond a weak conclusion to the effect that the speaker doesn't *know* the proposition expressed by 'Must Q'. But when we look at cases where ignorance is a salient possibility, it looks like the weak inference may be all we get. For example, it seems fine for a speaker who isn't sure whether Clark Kent is Superman to say 'If Clark looked similar to Superman, lots of people would think he was Superman', even though for all they know the antecedent is metaphysically necessary (and hence presumably true in all accessible worlds). So, despite the aesthetic appeal of a view that gives the two inferences the same status, we will refrain from positing a semantic presupposition of antecedent diversity (though we will also not say anything that depends on there *not* being such a presupposition).

1.4 Quasi-validity and materialism

In this section, we will show how associating indicative conditionals with the same context-sensitive accessibility parameter that features in the truth-conditions of epistemic modals can help to explain some puzzling features of indicative conditionals which have often been used to motivate the view that such conditionals have the same truth conditions (in all contexts) as the corresponding material conditionals.

Considering 'arguments' as consisting of a set of declarative sentences called 'premises' and another declarative sentence called a 'conclusion', we have been counting an argument as *valid* just in case, on any uniform resolution of context-sensitivity, the propositions expressed by the premises entail the proposition expressed by the conclusion. Let's say that an argument is *quasi-valid* iff its *premise-modalisation* is valid, where the premise-modalisation of an argument is the argument derived from it by prefixing 'It must be that. . . ' to each premise. Quasi-valid arguments tend to feel intuitively like excellent deductive arguments, even when they are not valid. For example:

(14) Either this is a horse or it's a donkey

It's not a horse.

Therefore it must be a donkey.

Even someone with an excellent logical training might naturally and unreflectively answer 'yes' when asked whether this is valid. Nevertheless, it is merely quasi-valid, and not valid. Clearly, if 'It must be a donkey' is in

the business of expressing a proposition at all, there are possibilities where the propositions that it is eligible to express are false even though those expressed by 'It is a donkey' (on the same interpretation of the pronoun) is true. And since being a donkey is incompatible with being a horse, any such possibility will be a counterexample to the validity of (14).[12]

A full account of epistemic modality needs to explain what's so good about merely quasi-valid arguments like (14), and why this kind of goodness is so readily mistaken for validity. Plausibly, the answers to these questions will involve fleshing out in some way or other the kinds of ideas sometimes discussed under the heading of 'the knowledge norm of assertion'—the thought being that in uttering a declarative sentence assertively, one is in some sense committed not merely to the truth of the proposition semantically expressed, but to its being known, or meeting some contextually flexible evidential standard that's also picked up by words like 'must'. One might also want to make a connection to the idea (R. C. Stalnaker 1978) that the characteristic function of an assertion is to add a proposition to the "common ground" of a conversation, and that the presence of the proposition expressed by *P* in the common ground is at least a sufficient condition for the truth of 'It must be that *P*'.[13] But for present purposes it doesn't much matter how exactly we nail these ideas down.[14]

One prominent argument for the 'materialist' view of indicative conditionals, according to which the indicative 'If *P*, *Q*' and 'Either not-*P* or *Q*' express necessarily equivalent propositions in all contexts, appeals to the seeming excellence of instances of the following argument-schema:

[12]Many authors have proposed that 'must' sentences do not express propositions, and extend the meaning of 'valid' in such a way that some arguments including non-proposition-expressing sentences, including (14), count as 'valid'. We will have more to say about these views in chapter 5; for now, we just want to point out that the mere fact that these views enable one to say that arguments like (14) are 'valid' is not by itself a reason to accept them, since those who think that 'must' sentences express propositions can say the same thing just by identifying validity not with truth-preservingness but with the status we are calling quasi-validity, i.e. truth-preservingness of the argument got by prefixing the premises with 'must'.

[13]Cite: Mandelkern,. . .

[14]A completely different (but compatible) kind of story about the goodness of quasi-valid arguments appeals to the idea that epistemic accessibility is often constrained by questions under discussion. The thought would be that following the assertion of some premises, some questions which those premises answer will always be salient in the relevant way, so that by default the accessibility parameter for epistemic modals and indicative conditionals will be resolved in a way that requires match with regard to the truth values of the premises; on this resolution of the parameter, the quasi-valid argument is in fact valid, since each premise entails its modalisation. While this idea has promise, we would need to handle it with care since (for reasons that will emerge more clearly in section 1.5) we think of constraint by questions as an often-natural option rather than any kind of strong default.

OR-TO-IF *P* or *Q*. Therefore if not-*P*, *Q*.

As many authors have noted, instances of this form just "feel valid": Stalnakert (1975) gives the example 'Either the butler or the gardener did it. Therefore if the butler didn't do it, the gardener did.' But if we took this appearance at face value, it would be a short step to materialism. (Given the logical equivalence of 'not-not-*P*' and *P* and the principle of Substitution in the Antecedent, the validity of Or-to-if guarantees that of the schema 'Not-*P* or *Q*; therefore if *P*, *Q*'. In the other direction, suppose that 'if *P*, *Q*' is true. 'Not-*P* or *P*' is true by the Law of Excluded Middle, so by Modus Ponens and Proof By Cases, 'Not-*P* or *Q*' is also true.) Thus if we want to deny the logical equivalence of 'If *P*, *Q*' and 'Either not-*P* or *Q*', we had better deny the validity of OR-TO-IF.

Similar remarks apply to the following schema

UNCONDITIONAL AGGLOMERATION *Q*. So if *P*, *P* and *Q*.

Instances of this form—for example 'He is an idiot; so if he is rich, he is a rich idiot'—have a similar feeling of validity to them. But again, the view that the schema is really valid leads to materialism under minimal additional assumptions.[15]

Given that quasi-valid arguments as well as valid ones can be expected to strike us as excellent, our view provides a ready rejoinder to these arguments for materialism. For since it is part of the view MUST-IF ('It must be that *Q*, so if *P*, *Q*') is valid, along with *Identity* and *Deduction in the Consequent*, both of the following argument-schemas are also validated:

MODALISED OR-TO-IF It must be that (*P* or *Q*). So if not *P*, *Q*.

MODALISED UNCONDITIONAL AGGLOMERATION It must be that *Q*. So if *P*, *P* and *Q*.

[15]Suppose that the material conditional 'Either not-*P* or *Q*' is true; then by UNCONDITIONAL AGGLOMERATION, 'If *P*, *P* and (either not-*P* or *Q*)' is true; so by *Deduction in the Consequent*, 'If *P*, *Q*' is true too. For the argument from 'If *P*, *Q*' to 'Either not-*P* or *Q*', see the previous paragraph.

Hence both OR-TO-IF and UNCONDITIONAL AGGLOMERATION are quasi-valid, which puts them on the same footing as the intuitively excellent argument (14) above. We don't see any pre-theoretic pressure to think that these arguments are any better than the likes of (14).[16]

(Note that the validity of MODALISED OR-TO-IF and MODALISED UNCONDITIONAL AGGLOMERATION depends on the decision to let the conditional be true when there is no accessible world in which the antecedent is true. If we had instead gone for a view on which conditionals are false in this case, we would need to say something more complicated and less satisfying about the status of OR-TO-IF and UNCONDITIONAL AGGLOMERATION.)

One way to make a positive case against the materialist view that the relevant arguments are in fact valid is to see what happens when premises and conclusions are embedded under operators like 'For all I know'. When the argument from P to Q is really valid (and easily recognised as such), 'For all I know P, so for all I know Q' should sound like a good piece of reasoning. But speeches like 'For all I know he is at the match, so for all I know if he had a car crash this morning he had a car crash and is at the match' and 'For all I know, it's true that she's either in Paris or on Mars (because for all I know she's in Paris); so for all I know, if she's not in Paris she's on Mars' seem bad. This is good prima facie evidence that the felt goodness of the relevant inferences should be explained by some feature other than straightforward validity.

(A related argument for materialism appeals not to intuitions about the validity of *arguments*, but to intuitions about the logical truth of individual sentences of the form 'If P or Q, then if not-P, Q'. The notion of quasi-validity does not provide a response to this form of argument, since there is no corresponding notion of 'quasi-logical-truth' distinct from logical truth proper. One can certainly allow for arguments with zero premises, and

[16]Our view also provides the resources to describe a way in which OR-TO-IF is better than UNCONDITIONAL AGGLOMERATION. Although MODALISED UNCONDITIONAL AGGLOMERATION is valid, asserting its premise does not in any way imply that the *presupposition* of its conclusion is true. In this respect it is like 'He doesn't regret anything, so he doesn't regret that he cut off his arm' or 'Everything in my garden is green, so every dog in my garden is green'. By contrast, asserting a disjunction carries the implicature that both disjuncts are epistemically possible for the speaker (readily explicable in neo-Gricean fashion), and this presumably also holds for 'It must be that (P or Q)'. So while the mere truth of the premise of MODALISED OR-TO-IF does not guarantee the truth of the presupposition of its conclusion, the implicatures carried by an assertion of the premise do guarantee this, at least in a standard context where epistemic possibility for the speaker suffices for accessibility.

identify logical truth with validity of the corresponding zero-premise argument; but quasi-validity coincides with validity for zero-premise arguments, since the result of prefixing 'must' to every member of the empty set is still the empty set. We think there is a good response to this argument too, but it requires some general points about conditionals which embed other conditionals which we will take up later. For the present, just note that it would be difficult to endorse this as an argument for materialism unless one were also willing to endorse the apparent logical truth of 'If this is either a horse or a donkey and it's not a horse, then it must be a donkey' as an argument for the view that 'It must be that P' expresses a true proposition whenever P does.)

It is worth comparing our response to the validity-theoretic arguments for materialism with a somewhat similar response offered by Stalnakert (1975). Stalnaker's view can be formulated in a way that fits the template of CLOSEST, and he agrees with us that a certain notion of epistemic possibility plays a distinctive role in the semantics of indicative conditionals. (Stalnaker takes the worlds that are epistemically possible in the relevant sense to be those that belong to the 'context set' of the relevant conversation, i.e. are consistent with everything "commonly presupposed" by those involved.) For Stalnaker, however, the distinctive role consists not in the exclusion of epistemically impossible worlds from the domain of accessibility, but a distinctive kind of closeness ordering. His proposal is that whenever w_1 and w_2 are in the context set of a certain conversation (taking place at the actual world, which may be distinct from both w_1 and w_2), and w_3 is not in the context set of that conversation, w_2 is closer to w_1 than w_3 is according to the closeness relation operative in that conversation. There is no special connection between epistemic possibility and accessibility—in fact Stalnaker suggests that every metaphysically possible world always counts as accessible, so that 'If P, Q and not Q' is true in a context only if P expresses a metaphysical impossibility in that context. Thus, on Stalnaker's view, MUST-IF is *not* valid, even if we stipulate an interpretation for 'must' on which the truth of 'It must be that P' requires P to be true throughout the context set.[17] So long as the truth of 'Must P' does not require that P is true in all accessible worlds (which it certainly does not if all metaphysically

[17]This is not a very plausible theory of 'must'—it struggles with the fact that it's fine to say 'It must be raining outside' when the speaker can see people walking in with umbrellas, but the audience cannot see them. Mandelkern (***) suggests a view where what matters for the truth of 'It must be that P' is that P be true in all worlds in the "prospective" context set.

possible worlds are accessible), any case where 'Must P' is true while the closest not-P world is a not-Q world will be one where 'Must (P or Q)' is true and 'If not-P, Q' is false.

Stalnaker thus cannot agree with us that OR-TO-IF is quasi-valid. Nevertheless, as his view does accord OR-TO-IF the status of a "reasonable inference", defined as one where 'every context in which the premises could appropriately be asserted or explicitly supposed, and in which it is accepted, is a context which entails the proposition expressed by the corresponding conclusion' (Stalnakert 1975). The crucial extra ingredient needed to derive this is the Gricean thought that asserting a disjunction 'P or Q' carries the implicature that both not-P and not-Q are epistemically possible (which for Stalnaker requires that true at some worlds in the context-set). Thus, any case where 'P or Q' is both appropriately asserted and accepted will be one where 'P or Q' is true at every world in the context-set, while both 'not P' and 'not Q' are true at some worlds in the context-set. So in any such case, 'if not-P, Q' will be true at every world in the context set, since the closest P world to each of these worlds is one of the 'not-P'-worlds in the context set, all of which are Q-worlds.

A similar implicature-based explanation of the positive status of OR-TO-IF will be available to anyone who regards the following argument as valid:

MUST-MIGHT-IF Must Q. Might P. Therefore, if P, Q.

So long as this is valid, OR-TO-IF will have whatever good status is shared both by the inference from P to 'Must P' and by the inference from 'P or Q' to 'Might P and might Q'.[18]

[18]Indeed, Stalnaker could agree that MUST-MIGHT-IF is valid if he held that 'Must P' requires P to be true throughout the context set, that 'Might P' requires P to be true somewhere in the context set, and that the actual world is always in the context set. His actual view, however, is that the actual world need not be in the context set, so there is no obvious way for him to agree that MUST-MIGHT-IF as valid. True, if he treated the truth of P throughout the context set as a sufficient as well as necessary condition for the truth of 'must P', he would regard the variant argument form derived from MUST-MIGHT-IF by prefixing the conclusion with 'must' as valid; however, insofar as there are reasons to think that 'must P' implies P, this theory of 'must' especially unpromising given Stalnaker's overall view.

Stalnaker could validate MUST-MIGHT-IF by adopting, instead of or in addition to the constraint on closeness already discussed, a constraint according to which all worlds in the context set have to be closer *to the actual world* than any world not in the context set. But he had a good reason for not adopting that constraint, namely that it is incompatible with the claim that no world is closer to any world than that world itself, without which there will be counterexamples to modus ponens.

Unfortunately, implicature-based accounts of the positive status of OR-TO-IF will not carry over to our other good-seeming argument-form, UNCONDITIONAL AGGLOMERATION. On Stalnaker's view, the premise Q could be true, and true at every world in the context set, even though the conclusion 'If P, P and Q' is false. This will happen when P is true at no world in the context set and the closest P world is a not-Q world. This is a major limitation in Stalnaker's diagnosis, since as we noted above, instances of UNCONDITIONAL AGGLOMERATION feel intuitively on a par with instances of OR-TO-IF.

A different tack that Stalnaker could take with UNCONDITIONAL AGGLOMERATION is to appeal to the presupposition of validity. Typically, when we hear someone utter the conclusion 'If Q, P and Q', we will start taking it for granted (if we weren't already taking it for granted) that Q is an epistemic possibility—this process of "presupposition accommodation" is one of the characteristic communicative functions of presupposition-carrying sentences. But given this presupposition and the epistemic necessity of the premise P, the conclusion must be true. To make this precise, say that an argument is *Strawson-valid* just in case on every uniform interpretation, the proposition expressed by its conclusion is entailed by the conjunction of all the propositions expressed by the premises, presupposed by the premises, or presupposed by the conclusion.[19] And say that an argument is *Strawson-quasivalid* just in case its premise-modalisation is Strawson-valid. The observation, then, is that since 'If Q, P' presupposes what 'Might Q' expresses, the validity of 'Must P; might P; therefore if Q, P and Q' guarantees the Strawson-quasivalidity of UNCONDITIONAL AGGLOMERATION.

Unfortunately, this status does not seem to be enough to account for the intuitive goodness of instances of UNCONDITIONAL AGGLOMERATION, since many arguments that are Strawson-valid (and hence a *fortiori* Strawson-quasivalid) don't seem intuitively that great at all. Consider for example:

(15) a. No-one will be smoking in the year 2050; therefore Fred will have stopped smoking by 2050

b. John regrets everything he has ever done; therefore John regrets drinking petrol

c. Every creature in the room is purring; therefore, the elephant in the room is purring

Given the standard view that the conclusions of these arguments presuppose, respectively, the propositions expressed by 'Fred will have been a smoker

[19]See von Fintel 1999.

before 2050', 'John killed his mother', and 'There is an elephant in the room', these arguments are all Strawson-valid. So *prima facie*, our view does a better job than Stalnaker's at explaining away the appearances of validity that drive the appeal of materialism.

One might worry that in going as far as we are going to vindicate the validity-judgments that motivate materialism, we will also be saddling ourselves with some of the classic problems of materialism—specifically, the so-called "paradoxes of material implication", arguments that are valid according to materialism but don't seem intuitively good at all. (*cite*) Here is the first of them:

FIRST "PARADOX" Q. So if P, Q.

On our view, this is quasi-valid.[20] We admit that it would be odd and suspicious if someone actually produced such an argument in the course of any piece of reasoning. But we don't think this is a good reason to deny that such arguments are quasi-valid: for 'This is a horse; so this must be a horse' is uncontroversially quasi-valid but also seems quite bizarre. If someone were to utter this sequence of sentences, we would feel pragmatic pressure to interpret the 'must' in some way that would give the conclusion some conversational point, and this will require some interpretation that doesn't tie the 'must' in a flat-footed way to what has been established at that point in the conversation.[21] In general, when we are dealing with super-simple arguments, intuitions of excellence will be confounded by the fact that such arguments have so little conversational point that they send us looking for non-obvious interpretations under which they are more tendentious or informative. Notice once we turn to only slightly more complex forms like UNCONDITIONAL AGGLOMERATION, which clearly stands or falls with the First Paradox, the intuitions of excellence start to fall into place. (Or in a similar vein, consider: 'John is in the room; so if Jill is in the room, two people are in the room'.)

The other classic objection to materialism turns on the following inference scheme:

SECOND "PARADOX" Not P. So if P, Q.

Again, this is valid on the materialist view and quasi-valid on our view, and it sounds like a truly awful template for argumentation. But note that

[20]Similarly, Stalnaker's view makes this argument form Strawson-quasi-valid.

[21]*cite Mandelkern on the need for a salient argument*

given the presupposition of non-vacuity, the argument 'It must be that not-P, therefore if P, Q' has the bad-making feature that if the premise is true, the presupposition of the conclusion is violated. In this respect it is similar to arguments like 'She hasn't stolen from anyone; therefore she doesn't regret stealing from her employer' (assuming a standard view about the presuppositional behaviour of 'regret'); 'No-one met anyone; therefore no-one met the King of France' (assuming a standard view about the presuppositional behaviour of 'the'); or 'Everyone is having a good time; therefore everyone who isn't having a good time is a terrorist' (assuming a somewhat more controversial view about the presuppositional behaviour of 'everyone'). The latter seems especially apt as a model for arguments involving conditionals. If someone actually produced such an argument, we would naturally reach for an interpretation on which the domain of the quantifier in the conclusion is wider than the domain of the quantifier in the premise, since there is clearly a strong tendency to interpret the quantifier in 'every F is G' in such a way that the truth of 'something is F' can be taken for granted.[22] And if we do widen the domain of quantification in this way, the proposition expressed by the conclusion will not be entailed by the proposition expressed by the premise. This seems like a reasonable way of vindication of our intuitive sense that there's something wrong with the Second Paradox as an argument-template, and driving a wedge between this argument and the intuitively good OR-TO-IF and UNCONDITIONAL AGGLOMERATION.[23]

[22]There has been considerable disagreement as regards whether this effect is properly explained by saying that 'Every F is G' *presupposes* the falsity of 'Nothing is an F'. For an overview see Heim and Kratzer 1998: §6.8.2. The most widely cited opponents of the presuppositional view, Lappin and Reinhart (1988), are primarily motivated by the combination of orthodox logical views with which we agree—such as that 'Every F is F' is valid—with a trivalent theory of presupposition (equating presupposition-failure with being neither true nor false); given this theory, the presuppositional view requires denying that all sentences of the form 'Every F is F' are true, which seems to conflict with the claim that such sentences are valid. This argument against the presuppositional theory of 'every' has no force in the non-trivalent framework we are working in.

[23]The fact that intuitions of validity are sensitive to presupposition-theoretic facts in this way might suggest that what these intuitions generally track is either Strawson-validity (explained earlier) or "presuppositional validity", where an argument is presuppositionally valid just in case on every uniform interpretation, the conjunction of all the propositions expressed or presupposed by the premises entails the conjunction of the propositions expressed or presupposed by the conclusion. (In the framework where presupposition-failure is identified with truth-valuelessness, this corresponds to the conclusion being true at every world where all the premises are true.) We are skeptical: the argument 'She doesn't regret stealing from her mother; therefore she stole from someone' seems fishy despite being

The idea that sometimes intuitively excellent arguments are quasi-valid rather than valid can also be used to undercut a certain way of arging for "strictism", the view that the indicative conditional 'If P, Q' has the same truth conditions as the so-called "strict conditional" 'It must be that (either not-P or Q)'. Someone might argue for strictism on the grounds that the argument 'If P, Q; therefore it can't be that P and not-Q' seems excellent in a way that requires treating it as valid. (If it's valid, the argument 'If P, Q; therefore it must be that (not-P or Q)' surely is too.) Our view makes these arguments merely quasi-valid: since 'If P, Q' logically entails 'Not-P or Q', 'It must be that (if P, Q)' logically entails 'It must be that (not-P or Q)'.[24]

1.5 How common are material-like readings?

We have said why we are not convinced by the central argument for materialism. However, the foregoing discussion also suggests that the task of arguing *against* materialism is going to be somewhat delicate. Given that opponents of materialism deny that the indicative conditionals are entailed by the corresponding material conditionals, one might prima facie expect that they would be willing to provide counterexamples to this entailment, where providing a counterexample would involve making a speech of the form '$P \supset Q$, but it is not the case that if P, Q'. But our account suggests that speeches like this will generally be unacceptable. Since MUST-IF is valid, so is the inference from 'Not (if P, Q)' to 'Not must (not-P or Q)', i.e. 'It might be that P and not Q'. So the counterexample-offering sentence validly entails something of the form 'P but it might be the case that not-P'. And sentences of this form are generally quite odd.[25]

We don't want to overstate the problem posed by the relevant kind of oddity. Speeches of the form 'P and if not P, Q' are sometimes fine.[26] And one could perhaps in the same mood accept speeches of the form 'P and not (if not P then Q)'—for example, 'Oswald shot Kennedy, but it's not the case that if he didn't, no-one did'. But we still wouldn't want to lean too heavily on arguments against materialism based on such premises. Materialists might reasonably respond that we are especially prone to error in deploying the stilted locution 'it is not the case that'—note for instance that 'It's not the case that every unicorn that he owns is coming to the party, since he doesn't own any unicorns' sounds initially fairly appealing.[27]

In our own thinking about materialism, we have been particularly moved by a well-known argument having to do with degrees of confidence. Consider:

(16) I'm moderately confident that if they did go out they went to the movies, but I am even more confident that they didn't go out.

The psychological state reported by this sentence is one that could not be maintained by someone who was certain of the material conditional theory, and formed credences about propositions expressed by indicative conditionals in accordance with the dictates of that theory. This pattern of credence formation involves never being less confident in an indicative conditional

presuppositionally valid.

[24] Some other argument-forms that are valid according to materialists and strictists, but merely quasi-valid on our account (for present-perspective indicative conditionals), are: contraposition ('if P, Q; therefore if not-Q, not-P'), transitivity ('If P, Q; if Q, R; therefore if P, R'), and antecedent strengthening ('If P, R; therefore if P and Q, R').

[25] The task of explaining why they are odd is of a piece with the task of explaining what's good about the inference from P to 'must P', and about quasi-valid inferences in general. A first-pass explanation appeals to the knowledge norm of assertion: in asserting P one represents oneself as knowing that P, while 'might P' entails that one does not know that P, so the overall impact of the speech act is incoherent A more sophisticated explanation will have to take account of the context-sensitivity of 'must' and 'might': in some way we will not here attempt to fully work out, resolving the context sensitivity of modals in a way that makes it hard for 'must P' to be true also raises the bar for a flat-out assertion of P.

[26] It is interesting that such sentences don't sound as bad as 'P and might not-P', e.g. 'Oswald shot Kennedy but he may not have'. We suggest that this is because, although both sentences require something akin to non-uniform interpretation—a transition from the relatively demanding notion of epistemic possibility implicitly invoked by the first conjunct to the relatively lax notion explicitly invoked by the second conjunct—this kind of transition is more natural when it is driven by presuppositions than when it is required by expressed content. The fact that presuppositions are in the background, and are signaled as things to be taken for granted rather than up for debate, makes them especially useful as a way of guiding hearers towards non-uniform interpretations: contrast 'Everyone came to my party, and everyone who wasn't at the party but read about it in the newspapers was jealous' with 'Everyone came to my party, and some people weren't at the party but read about it in the newspapers'.

[27] Arguments against materialism that turn on conditionals embedded under quantifiers provide one way around the problem of assertability. For example, one might object that materialism entails that in a setting where exactly one of several candidates won the election, 'Some candidate lost if she won' is true, whereas in fact no candidate lost if she won. However, as we will be discussing in chapter **??**, materialists have some defensive resources involving domain restriction. More generally, sentences in which conditionals are embedded under quantifiers turn out to be challenging not just for materialism but for many competing views, so the dialectic concerning them will have to take the form of a careful comparative investigation.

than in the negation of its antecedent, since according to the material conditional theory, the conditional is entailed by the negation of its antecedent. But it seems obviously fine to be the kind of person of whom (16) is true. Similarly, it seems paradigmatically fine to be the sort of person that is 50% confident that if a certain fair coin was tossed it came up tails, but less than 100% confident that it was tossed, and thus less than 50% confident that it was tossed and came up tails. But again, since according to the material conditional theory the conditional is entailed by its consequent, forming credences in accordance with that theory would involve never being more confident in a conditional than in its consequent.

The point here is not just about sentences in which indicative conditionals occur embedded under 'confident that'. For one thing, it arises with many different kinds of embeddings, including those whose connection to the concept of confidence is not so straightforward. For example, suppose that we know that Smith was playing poker, and we now learn that his opponent is Fred, a mediocre but rich player who generally bets big. The following speech is natural in this setting:

(17) It's starting to look like Smith won big if he won.

But given Fred's mediocrity, the evidence that he was Smith's opponent may actually be evidence *against* the corresponding material conditional (*Smith either lost or won big*), in virtue of being evidence against the proposition that Smith lost; the acceptability of (17) is thus mysterious from the materialist point of view.

The pattern of materialism-unfriendly confidence judgments also shows up in cases where, instead of having a conditional embedded under an operator like 'confident', we have anaphoric reference back to the proposition expressed by a bare conditional earlier in the discourse. Consider:

(18) If they went out they went to the movies.

— You're probably right / I'm pretty confident of that / There's a good chance that they did / That's fairly likely to be true / . . .

It is easy to spell out the scenario in such a way that these replies are all in order, but where the second speaker is extremely confident that they didn't go out at all, and so is clearly not forming credences in accordance with the dictates of the material conditional theory.[28]

Similar points can be made about desire-like attitudes. Consider hope:

[28]These anaphoric cases are important, since some authors (Kratzer 1986, Rothschild forthcoming) have considered accounting for the behaviour of conditionals under operators like 'probably'—perhaps even including 'confident that'—by adopting a logical form for the 'confident' sentences which does not have the logical form of the bare conditional sentence as a constituent, and correspondingly denying that the truth-conditions of the 'confident' sentences involve the relevant subject's relations to the propositions that would be expressed by utternces of a bare conditional. von Fintel (MS; 2007) also notes the failure of such accounts to extend to examples like (18).

(19) I'm hoping that I've won a blue car if I've won a car

On the material conditional account, the proposition expressed by 'I've won a blue car if I've won a car' is equivalent to the one expressed by 'Either I've won a blue car or I haven't won a car'. But on the natural way of fleshing out the scenario, (20) would seem to be false:

(20) I'm hoping that I have either won a blue car or haven't won a car

Of course, it is open to materialists to admit that people do in fact form credences and pro-attitudes in the ways that we have described, but to put this down to their blindness to the truth of materialism. Perhaps, once one has become enlightened, one should simply revise one's credences (and hopes, etc), so that there will no longer be cases where one is less confident in a conditional than in its consequent or the negation of its antecedent.[29] This need not involve saying that ordinary folk are *irrational*, since failure to know the truth of a true theory, even in philosophy, need not be a failure of rationality. Nevertheless, there is generally a strong presumption against the truth of semantic theories that would have to induce large-scale revision of ordinary patterns of usage if taken to heart. Consider, by analogy, the theory that 'know' expresses the relation of truly believing: if taken to heart, this would recommend a substantial overhaul of a central pattern of judgments that have guided users of 'know' and its cognates since antiquity. This looks like a powerful objection to the view. Proponents will say things like 'My view is more simple, and we already know that ordinary people make mistakes all the time', but this does not seem adequate to the force of the objection.

The attitude-based objection to materialism would of course have little force if it turned out that no theory could underwrite the relevant patterns of attitudes. Indeed, there is a body of literature (going back to Lewis 1976 and R. C. Stalnaker 1976) that attempts to identify the relevant pattern, and then

[29]It is an open question whether the update involved in such enlightenment is appropriately modeled using standard conditionalisation.

argue that no way of associating propositions with conditionals can vindicate that pattern. We will discuss this properly in **??**. For now, we just want to remark, first, that some of the literature in question assumes the falsity of the kind of fine-grained context-sensitivity that we have been advocating in this chapter, and second, that the particular confidence-judgments about particular cases that we were relying on need not be significantly undermined by the difficulty of subsuming them under some completely general principle that applies to all indicative conditionals, whatever their logical structure.

The confidence-theoretic arguments suggest that very often, an indicative conditional expresses a proposition that is not known to have the same truth value as the corresponding material conditional, and that fairly often, the proposition expressed by the indicative conditional is actually false even when the antecedent is false or the consequent true. Of course there are also many occasions where an indicative conditional is used to express a proposition that is known to have the same truth value as the corresponding material conditional. Boringly: when the material conditional is known to be false, the corresponding indicative conditional can also be known to be false (since it entails the material conditional). Slightly less boringly: if the operative notion of accessibility is such that the material conditional is known to be true in all accessible worlds, then the indicative conditional can be known to be true (since its truth is entailed by the proposition that the material conditional is true in all accessible worlds.)[30] But the structure of CLOSEST also allows that in some contexts, the accessibility parameter may be set in such a way that we can know that the conditional has the same truth value as its material counterpart without knowing which truth value this is. For a conditional 'If P, Q', this will happen whenever the operative accessibility relation is such that no world compatible with what we know is a 'not-P'-world from which a 'P and not Q'-world is accessible. This condition guarantees that at every 'not P'-world compatible with what we know 'If P, Q' is true (and thus has the same truth value as '$P \supset Q$'). Meanwhile, the requirement that every world is the closest world to itself guarantees that 'If P, Q' always has the same truth value as '$P \supset Q$' at P-worlds (whether or not they are compatible with what we know).

The theory that accessibility is sometimes constrained by salient questions under discussion suggests a way in which this might happen. Suppose for example that after Smith left on a parachute jumping trip, we found

[30] If accessibility is just being consistent with what we know, this condition boils down to our knowing that we know the material conditional.

out that there were some non-working parachutes lying around the airport. The question whether the parachute Smith took with him was working is a salient and pressing one from our point of view. We think Smith is likely, but not certain, to notice a broken parachute and so is unlikely to jump to his death. As we investigate further, we find to our dismay that *most* of the parachutes were broken. Suppose one of us pessimistically utters 'If Smith jumped, he died'. There are two natural ways of reacting to this speech. On the one hand, we could say 'That's quite improbable—remember that Smith is a pretty experienced skydiver and is likely to have thoroughly checked his parachute before deciding whether to jump'. But on the other hand, faced with the evidence of the abundance of broken parachutes, we might say 'More likely than not you're right, but don't forget the possibility that he got one of the good parachutes'. Plausibly these two reactions correspond to two resolutions of the context-sensitivity of 'If Smith jumped, he died'. A natural diagnosis is in the second case, the accessibility relation requires match with respect to the salient question under discussion, namely whether Smith's parachute was working. Since our knowledge rules out the possibility that Smith died and had a working parachute, the possibility that Smith jumped without a working parachute and din't die, and the possibility that Smith had a working parachute and didn't jump, the effect of imposing this constraint on accessibility is to ensure that no worlds where Smith jumped and didn't die are accessible from any of the worlds compatible with our knowledge where Smith didn't jump (since all of these worlds are worlds where Smith didn't have a working parachute). By the observation in the previous paragraph, this is sufficient for the relevant reading of 'If Smith jumped, he died' to agree in truth value with the material conditional 'Smith jumped $\supset$ Smith died' at all epistemically possible worlds.

One might worry that the proposal that the accessibility parameter for indicative conditionals is readily constrained by questions under discussion would predict that it is much more common than it in fact seems to be for indicative conditionals to behave in a material-like way. After all, when someone utters 'If P, Q' the question whether P is true is always salient during the utterance, and often salient prior to it. But if the conditional is interpreted relative to an accessibility parameter requiring match with regard to whether P, it must agree in truth value with '$P \supset Q$' at every world, since there is no way for a P-and-not-Q world to be accessible from a not-P world. However, given the presupposition of nonvacuity, this particular kind of constraint by a question under discussion will not take be intended in normal circumstances:

if being accessible requires agreeing with actuality with regard to the truth value of P, the only way we could reasonably take it for granted that there are accessible P-worlds would be if we could reasonably take it for granted that P is true. But if we are taking it for granted that P is true we are unlikely to assert 'If P, Q', since we will then be in a position to assert something stronger and simpler (or equally simple) such as 'P and Q' (or 'Must Q', or perhaps just Q). Similarly, although the question whether Q is true is often under discussion in the setting of an utterance of 'If P, Q', the accessibility parameter is unlikely to be constrained by this question: this interpretation makes 'If P, Q' equivalent to the disjunction '(Must not P) or Q', which is in turn equivalent to Q modulo the presupposition of nonvacuity—but clearly there would be no point in uttering 'If P, Q' to convey a total content that could equally well have been conveyed by just uttering Q.

McDermott (1996) introduces some interesting examples where conjunctions of conditionals seem to have the simple truth-conditions that a material-conditional analysis would predict. Suppose that after a die has been rolled but before the result has been revealed, someone asserts (21):

(21) If the number showing is even it's 6, and if's odd it's 1.

(Maybe the speaker thinks the die was weighted, or is hoping to get a reputation for paranormal abilities. . . .) In this setting, (21) feels straightforwardly equivalent to the claim that the number showing is 1 or 6. In particular, if it turns out that the die landed on 6, we don't think along the following lines: the first conjunct is true, but the status of the conjunction is still unclear, because what we've learnt by looking at the die doesn't settle whether the number showing was odd if it was 1—in the terms of our analysis, it doesn't settle whether the die lands 1 at the closest accessible world where it lands on an odd number. The equivalence in question is smoothly explained by a view on which the conditionals are interpreted as material conditionals: on such a view, the second conjunct is automatically true in worlds where the die landed on an even number, and the first conjunct is automatically true in worlds it landed on an odd number.

There are many possible settings of the accessibility parameters for the conditionals in (21) that would make the conjunction equivalent to 'It was a 1 or a 6', some of which seem more natural than others. One that seems quite natural takes accessibility (for both conditionals) to require (in addition to epistemic liveness) match with regard to the very salient question whether the die landed on an even or odd number. This will make the conditionals necessarily equivalent to the corresponding material conditionals: for example, in the scenario where the die lands on 1, there are no accessible worlds where it lands on an even number, so the first conjunct is vacuously true (while the second is non-vacuously true).

An objection to the proposal that the relevant interpretation of (21) is constructed in this way is that it conflicts with the presupposition of non-vacuity. Of course, (21) is a conjunction of conditionals, not a conditional, so to extract any predictions for its presuppositions, we will need to appeal to some claims about "presupposition projection", the manner in which the semantic presuppositions of compound sentences are determined by those of their constituents. For the case of conjunction, there is some dispute in the literature about how this should work. Some examples are well explained by the simple theory that when P presupposes p and Q presupposes q, 'P and Q' presupposes the conjunction of p and q. (For example, 'Her house is expensive and her car is cheap' seems to presuppose what 'She has a house and she has a car' expresses.) But there are other well-known examples are problematic for this theory, and suggest the following more complicated theory: when P expresses p and presupposes p' and Q expresses q and presupposes q', 'P and Q' presupposes $p' \wedge ((p \wedge p') \supset q')$, the conjunction of p with a material conditional whose consequent is q' and whose antecedent is the conjunction of p and p'.[31] For example, 'He has a car that he keeps in his garage and his car is expensive' does not seem to presuppose what 'He has a garage and he has a car' expresses, but only what 'He has a garage' expresses—which is logically equivalent to 'He has a garage $\wedge$ ((he has a garage $\wedge$ he has a car that he keeps in his garage) $\supset$ he has a car)'. The question how to reconcile this impasse (the 'proviso problem') is one of the major current debates in the theory of presuppositions (**cites**). But for our purposes, it suffices to observe that on both the simple theory and the more complicated theory, conjunctions always inherit the presuppositions of their *first* conjunct. So we would predict that (21) will semantically presuppose that its first conjunct isn't vacuously true, which on an interpretation where accessibilityrequires match with regard to whether the die lands on an odd number would require that the die lands on an even number. But clearly the utterance of (21) carries no suggestion that we can take it for granted that the die lands on an even number—rather the contrary. Moreover, even on the more complicated theory, we would also predict that due to its second

[31] While the standard version of this story uses a material conditional, other interpretations of the conditional would do just as well with the data motivating the theory.

conjunct (21) would presuppose the material conditional whose antecedent is the conjunction of the propositions expressed and presupposed by the first conjunct, and whose consequent is the presupposition of 'if the number showing is odd it's 1', which on the proposed accessibility relation entails that the number showing is odd. So on either theory, it will be impossible for (21) to have the status 'true with only true presuppositions' on the proposed interpretation.

How big a problem this is for the proposal depends on how willing we are to think of the association between semantic presuppositions and actual speaker commitments as a contextually defeasible matter rather than something as robust as the association between semantically expressed content and speaker commitments. As noted in 1.3, some of the phenomena that get standardly lumped together under the label 'presupposition', including the tendencies for hearers to infer from 'He doesn't know *P*' to *P* and from 'I have not stopped smoking' to 'I used to smoke', do in fact seem rather delicate and easily defeated. Moreover, interestingly, there are examples which suggest that one possible defeating factor is the utterance of a conjunctive sentence whose conjuncts have conflicting presuppositions: sometimes, conjunctions for which the standard theories of presupposition projection for conjunctions would predict a contradictory or otherwise false presupposition seem perfectly fine. For example

(22) He doesn't know she is guilty, and he doesn't know she is innocent

would be standardly predicted to presuppose either the contradictory proposition expressed by 'She is guilty and she is innocent', or something equivalent to the proposition expressed by 'She is guilty and he knows she is guilty' which contradicts the asserted content, whereas in fact it is perfectly fine and indeed seems to carry no suggestions in particular that anything is to be taken for granted (other than the genders of the referents of the pronouns). Similarly,

(23) I won't salute the king and I won't salute the regent

would be standardly predicted to presuppose that there is a king; but if it's taken for granted that if there's a king there isn't a regent and vice versa, an utterer of (23) will probably not be taken to be committed to the presupposition predicted by standard lore.[32]

[32]We had better not overgeneralise the lesson here: sometimes, conjunctive sentences for which we would predict presuppositions which are contradictory, or which contradict their assertive content, really sound terrible, and we still want to be able to appeal to their presuppositions to explain why they are terrible. For example, we want to be able to say that 'If *P*, *Q* and if *P*, not *Q*' is bad because the only way for it to be true is for it to have a false presupposition. (***Is there an interesting contrast to be drawn here between sentences whose presupposition is contradictory by itself and sentences whose presupposition merely contradicts their assertive content?)

The presupposition of nonvacuity seems to sometimes disappear in similar ways in conjunctions. Suppose that in a poker game, I know that Sally's four cards are either the ten, jack, queen, and king of spades, or two nines and two eights. I have two kings and three aces. Sally folds; I look at the top card and see that it is the nine of spades. I say

(24) If she had drawn that card and had a straight she would have won, but if she had drawn that card and had a full house she would have lost.

This seems acceptable given what I know. To make it acceptable, it seems that we need an accessibility parameter that holds fixed the actual makeup of Sally's first four cards and my five—otherwise, it's puzzling how my actual knowledge concerning the values of these cards would license the assertion of (24). However, if the accessible worlds must match actuality in these respects, inevitably one of the two counterfactuals will be vacuously true. Given these precedents, it is not unreasonable to diagnose (21) as another case of a similar sort. After all, interpreted using the vacuity-friendly accessibility relation, it is just like (22) in that the presuppositions of the two conjuncts are jointly inconsistent.

An alternative approach to (21) is to find some other values of the accessibility parameter—perhaps different for the two conjuncts—which gets the conjunction to be equivalent to 'The die is showing 1 or 6' in a way that doesn't require either conjunct to be vacuously true. For example, we could take the accessibility relation for both conditionals to be the intersection of epistemic accessibility with an two-cell equivalence relation in which worlds where the die landed 1 or 6 form one cell, and worlds where it landed 2–5 form the other cell. This also makes (21) true when the die landed 1 or 6 and false otherwise, and guarantees that neither conditional is vacuously true. For example, if it landed 1, the first conjunct is nonvacuously true (since the actual world is the closest accessible odd-number world), and the second conjunct is also nonvacuously true (since there are accessible worlds where an even number is showing, in all of which the number showing is 6). While this approach gets the desired truth-condition, the accessibility constraints

it requires will often be rather gerrymandered—e.g. to deal in a similar way with 'If Mary goes to Paris she'll have a good time and if she doesn't go she'll have a bad time' one will need to use a partition where worlds where Mary has a good time in Paris or a bad time elsewhere are in one cell, and worlds where Mary has a bad time in Paris or a good time elsewhere are in another. Moreover, the approach in terms of "presupposition cancellation" explains something mysterious, namely that even when we are in a mood to count an actual result of 1 as vindicating (21), we are reluctant to just assert the first conjunct on its own ('I was right on both counts—if the number showing was even it was 6'). This might be explained by as a matter of the presupposition retaining its grip on us when we consider the conjunct individually. By contrast, if the intended interpretation of (21) involved an accessibility parameter making both conjunts non-vacuously true, it is harder to see what would be holding us back from following up by asserting them individually.

[. . .]

1.6 Accessibility for counterfactuals

As many authors have observed, our standard for evaluating a counterfactual whose antecedent concerns a particular period of time involves helping ourselves to all kinds of facts about earlier times. For example, if we know John has had breakfast every day for the last year, we will unhesitatingly endorse

(25) If John had forgotten to have breakfast on Tuesday morning, that would have been the first time this year.

The phenomenon is pervasive: even when the consequent of a counterfactual isn't about the past, in deciding what to make of it we will typically be tacitly "holding the past fixed". For example, in reasoning about whether to accept

(26) If I had run a four minute mile this morning, I would have been extremely surprised.

we seem to be holding fixed the speaker's previous track record.

Our favoured account of this phenomenon appeals to fine-grained context-sensitivity in the accessibility parameter. When considering a counterfactual whose antecedent is about a certain period, it is natural to resolve its context-sensitivity in such a way that the accessible worlds are all required to match

with respect to earlier times.[33] Notice that the closeness relation plays no role in this account of why counterfactuals like (25) are true: *all* the accessible worlds where John forgets to have breakfast on Tuesday are worlds where it his his first time doing so this year. So for example, while our account suggests that (25′) is true in the context *it* naturally evokes for the same reason as (25), it provides no reason to expect (25′) to be true relative to context naturally evoked by (25):

(25′) If John had forgotten to have breakfast on Wednesday morning, that would have been the first time this year.

After all, the context evoked by (25) is one where the accessible worlds are merely required to match with respect to history up to Tuesday; since some of them involve forgetting to have breakfast on Tuesday and on Wednesday, the accessibility facts provide no guarantee for the truth of (25′) in this context. And indeed, the account of closeness that we will develop in chapter 2 will provide no grounds for confidence that the closest worlds where John forgets breakfast on Wednesday are not worlds where he forgets breakfast on Tuesday as well. Meanwhile, while (25″) is non-vacuously true in the context it naturally evokes, it is vacuously true in the context evoked by (25).

(25″) If John had forgotten to have breakfast on Monday morning, that would have been the first time this year.

By contrast, Lewis's influential treatment of counterfactuals tells a rather different story about why holding the past fixed is a reliable way of evaluating counterfactuals. For Lewis, what does the work is closeness, not accessibility—indeed, his account would work even on the assumption that all metaphysically possible worlds are accessible in all contexts. Moreover, Lewis's account of the phenomenon does not rely in any essential way on context-sensitivity. For example, he thinks that there is a single context on which we can knowledgeably utter (25), (25′), and (25″): given that we know that John in fact had breakfast every day this year, we know that worlds where John skips breakfast for the first time on Wednesday are closer than worlds where he skips breakfast for the first time on Tuesday, which

[33] The match may not be perfect (cite), and it may need to be restricted to allow for a smooth transition into the kind of event described by the antecedent. Indeed for certain antecedents — consider 'If a giant comet had struck Washington DC yesterday afternoon. . . .' — we may need quite a long stretch of preceding time where there is nothing like exact match in order to achieve the smooth transition.

are in turn closer than worlds where he skipped breakfast on Monday. More generally, the closeness relation that Lewis thinks is standard is one on which worlds whose history diverges from that of the actual world at later times are ipso facto closer than worlds where history diverges earlier (Lewis 1979).

This feature of Lewis's view leads to some well-known oddities that our view avoids. John Pollock (reported in Nute 1980) noticed that Lewis's view seems to underwrite counterfactuals such as the following:

(27) If my coat had been stolen last year it would have been stolen on December 31st.

Given that my coat was not in fact stolen, on Lewis's view worlds where it is stolen are closer to the extent that they diverge later from the actual world.But intuitively, unless I have some special reason to think that my coat was unusually vulnerable to theft on December 31st, I have no right to assert (21).

It would really be quite bizarre if we had to start computing counterfactuals in the way seemingly recommended by Lewis. Just to take one more example: suppose that, sadly, someone fell from a ship and drowned while people stood and watched without anyone diving in to help. On Lewis's approach, there seems to be a quite strong reason to accept

(28) If someone had dived in to try to help her, she would still have drowned.

After all, if the later the divergence the closer, then the closest worlds will be ones where someone dives in too late, even if there are plenty of worlds where a more timely rescue attempt would have saved her.

The examples cry out for a treatment on which the only period of history that is completely held fixed is one that predates the period in question—last year in the case of (21), the period during which the drowning victim was in the water in the case of (28). A contextualist approach like ours can readily accommodate this. Given the antecedent of (21), it would be completely unmotivated to select some time after the beginning of last year as the basis for resolving accessibility: the natural constraint on accessibility will allow more or less indiscriminate importation of facts about the actual world only for times prior to last year.

One possible response to the problems for Lewis's account would be to maintain that for some reason (21) is computed as equivalent to

(21′) On any day last year, if my coat had been stolen on that day it would have been stolen on December 31st.

It's not obvious what would motivate this seemingly *ad hoc* mechanism beyond a desire to avoid the counterexamples. But in any case, the proposal is clearly inadequate. Suppose that halfway through the year I moved from a dangerous neighbourhood to a much safer one; then I could naturally assert

(29) Probably if my coat had been stolen last year, it would have been stolen during the first half of the year.

By contrast, (29′) is obviously unassertable:

(29′) Probably on any day last year, if my coat had been stolen on that year it would have been stolen during the first half of the year.

The problems raised by the coat and drowning examples are not peculiar to the details of Lewis's approach, or even to the possible worlds framework. Given certain minimal logical assumptions, problems of this sort will arise for any view that attempts to secure the reliability of 'holding the past fixed' for antecedents about different times without appealing to context-shift. Suppose that there was a thunderstorm all day Saturday in the area that destroyed many trees, but my tree luckily survived. Now consider

(30) If this tree had been in pieces on Sunday morning, I would have been too upset to have breakfast.

This is a typical example of the kind of sentence that is evaluated by holding the past fixed; so it's plausible that in the context naturally evoked by (30), (31) is true:

(31) If the tree had been in pieces on Sunday morning, it would have been intact on Saturday at bedtime.

(Let's take it that I am not so devoted to the tree that I would be at the breakfast-refusing level of sadness if I had already known of its destruction going to sleep.) But surely (32) and (33) are true as well (in the relevant context):

(32) If the tree had been in pieces on Sunday morning, it would have been destroyed on Saturday or earlier.

(33) If the tree had been destroyed on Saturday or earlier, it would have been in pieces on Sunday morning.

Denying either of these (32) looks completely unpromising—trees get to be in pieces by being destroyed, and they don't subsequently reform. But given (32), (33), and (31), we are in a position to apply the inference-schema sometimes called 'CSO', according to which arguments of the following form are valid:

CSO If *A* then *B*. If *B* then *A*. If *A* then *C*. So, if *B* then *C*.

Plugging in 'The tree was in pieces on Sunday morning', 'The tree was destroyed on Saturday or earlier', and 'The tree was intact on Saturday at bedtime' for *A*, *B*, and *C*, we can derive

(34) If the tree had been destroyed on Saturday or earlier, it would have been intact on Saturday at bedtime.

But (34) seems intuitvely problematic in the same way as the coat and drowning examples considered earlier.

CSO is valid according the best-known logics for counterfactual conditionals, namely those of Stalnaker and Lewis. And to our my minds its popularity is well-deserved. While it may not be immediately compelling when first encountered, there are several argumentative routes to it from principles whose appeal is more immediate. One route consists of

RT If *A* then *B*. If *A* and *B* then *C* . So, if *A* then *C*.

RCV If *A* then *B* and *C*So, if *A* and *B* then *C*.

(These patterns are sometimes called 'Cumulative Transitivity' and 'Very Limited Antecedent Strengthening, respectively'.) For suppose the premises of CSO hold: if *A*, *B*; if *B*, *A*; and if *A*, *C*. Given the first and third, by RCV it follows that if *A* and *B* then *C*; and from this and the second premise (If *B*, *A*) RT yields that if *B*, *C*.[34]

Reliance on RT seems to pervade a great deal of our ordinary reasoning using conditionals, both indicative and counterfactual. Consider how good the following piece of reasoning looks: 'If they attacked they would use their cavalry; if they attacked using their cavalry, they would win; so if they attacked, they would win'. And consider how terrible the following speech sounds: 'If he had gone to the party, he would have gone with Janet, and if he had gone to the party with Janet, he would have had a good time, but I doubt that if he had gone to the party he would have had a good time'. Clearly there is some general principle underlying the goodness of the good reasoning, and the badness of the bad speech, and it is hard to see what it could be if RT is invalid. Regarding RCV, similar remarks apply to the following excellent argument (which involves a contraposed application of RCV): 'If they attacked they would use their mercenaries; but it's not the case that if they attacked using their mercenaries they would win; so it's not the case that if they attacked they would win', and also to the following appalling speech: 'If he had gone to the party he would have gone with Janet, and had a great time; however if he had gone to the party with Janet, he would have had a terrible time.' (Note too that all of this looks equally compelling if we shift everything to an indicative form: 'If they attacked they used their cavalry', etc.)

[. . .]

Given the general considerations for thinking that context-sensitivity is pervasive, and the costs of giving up CSO, it seems to us much more appealing to explain the relevant data by appeal to context sensitivity rather than jettisoning CSO. On the picture we favour, (34) is indeed straightforwardly true in the context naturally evoked by (30), but would be a very risky assertion in the context it naturally evokes. Those who endorse CSO but think that there is no relevant context-sensitivity in play will, like Lewis, be forced to accept the problematic (34).

The diagnosis of context-sensitivity enjoys substantial independent support, since there are many cases where counterfactuals with the same antecedent suggest different ways of holding the past fixed. For example, each of (35a) and (35b) seems to be true in the natural context they suggest:

(35) a. If I had been in the Caribbean this morning I would have been feeling refreshed

b. If I had been in the Caribbean this morning I would have been exhausted from travelling

In these cases it is the consequent rather than the antecedent that provides the linguistic cues as regards which portions of the past to hold fixed. (Of course there may be non-linguistic cues as well.) On a view where (35a) and

[34] Also note that CSO entails RT and RCV, and that the two are equivalent modulo CEM which we will defend in chapter 2.

(35b) are both true in the same context, one would have to either live with bizarre counterfactuals like

(36) If I had been in the Caribbean this morning I would have been feeling refreshed and exhausted from travelling

or else postulate widespread violations of an even more obviously compelling inference rule, namely CC ('Finite Agglomeration'):

CC If *A*, *B*. If *A*, *C*. So if *A*, *B* and *C*.

The appeal to context-sensitivity is thus pretty much inevitable when it comes to sentences like (35a) and (35b). We see no reason to deny that the phenomenon manifested by the previously examples is any different: in both cases, contextual triggers determine how much of the past is held fixed in all accessible worlds.

Lewis, for his part, is willing to play the context-sensitivity card in a much more limited way. His picture is that there is a "standard" resolution of context-sensitivity under which worlds that diverge later are closer, and a separate category of "backtracking" resolutions of context-sensitivity where a different standard of closeness is in play. Lewis gives various examples where he thinks a backtracking resolution of context-sensitivity is natural, such as 'If Jim were to ask Jack for help today, there would have to have been no quarrel yesterday' (Lewis 1979, p. 33). He might say that one or both of our sentences (35a) and (35b) fall into this category as well.[35] Lewis never develops a theory about what plays the role of the closeness relation in backtracking contexts. But we think there is very little prospect for a theory that posits a *single* closeness relation for backtracking contexts. Typically, even in evaluating a counterfactual in a "backtracking" way we still draw freely on quite a lot of facts about the past—for example, in evaluating (35a) we probably draw on the fact that in the not-so-distant past the speaker was not feeling refreshed. There is no prospect of a general rule saying how much of the past to hold fixed for evaluating backtrackers. Given that fine-grained context-sensitivity is thus hard to avoid when dealing with the counterfactuals Lewis categorises as backtrackers, it would not at all

[35]While the examples Lewis focuses on are sentences where the consequent directly concerns times earlier than those mentioned in the antecedent, he is careful to leave open the possibility that the non-standard resolutions of context-sensitivity required to handle those are also natural for a range of other counterfactual utterances that don't have this structure.

surprising if (as we contend) it also extends to those he would categorise as "standard".

As well as requiring some bullet-biting about cases like Pollock's coat, Lewis's approach also requires him to give up a prima facie compelling principle about the logic of counterfactuals. If time extends infinitely into the future, and worlds are closer the later they diverge from actuality, then for every non-actual world, there is a closer one. The closeness relation thus fails to obey the 'Limit Assumption', which says that every set of worlds contains some worlds that are as close to actuality as any world in the set. Lewis is well aware of this structural feature, and crafts his truth conditions accordingly: a counterfactual is true iff either there is no (accessible) world where the antecedent is true, or at least one world where antecedent is true is such that every world where the antecedent is true that is at least as close as it is one where the consequent is true. However, the upshot of this semantics is that counterfactuals do not generally obey the following principle:

AGGLOMERATION If some propositions each would have been true if a certain condition had obtained, then if that condition had obtained, all of them would have been true.

On Lewis's account AGGLOMERATION is not true in general, although it is true for finite collections of propositions. To see why it can fail in the infinite case, consider the collection of propositions which contains, for each time *t*, the proposition that history proceeds just as it actually does until *t*. Each conditional of the form

(37) If history had diverged from actuality at some point, then history would have proceeded just as it actually does until *t*.

will be true, since for any *t* worlds which diverge after *t* are closer than worlds which diverge earlier. But clearly (38) is false:

(38) If history had diverged from actuality at some point, then each *t* is such that history would have proceed just as it actually does until *t*.

As Pollock (1976*a*,*b*), Fine (2011), and others have noted, the intuitive force of AGGLOMERATION carries over very smoothly from the finite to the infinite case. It is thus a uncomfortable feature of Lewis's view that it requires driving a wedge between the two kinds of cases.

This feature is very hard to avoid on any account that attempts to accommodate the phenomena without appealing to context-sensitivity. For on such an account, all sentences of the form

(39) If things had gone differently on day n or later, things would have been the same up to day n

will be true in a single context. Moreover, given the abundance of causal influence in the futurewards direction, there is strong pressure to accept all conditionals of the form

(40) If things had gone differently on some day, things would have gone differently on some day later than day n.

After all, a scenario where the only differences occur before day n and then history then re-converges to that of the actual world seems quite far-fetched. But obviously (41) is also true irrespective of context:

(41) If things had gone differently on day n or later, things would have gone differently on some day.

So by CSO, we can conclude that all the conditionals of the following form are true in the same context:

(42) If things had either gone differently on some day, things would have been the same up to day n.

This is obviously inconsistent with AGGLOMERATION. The case for (40) is perhaps not quite watertight: one might think that for all we know, if there had been any divergence at all it would have been a very slight and temporary divergence confined to some initial finite period. But we can forestall this response by substituting the property of being a day such that things go differently from how they actually go on it and all subsequent days for the property of being a day on which things go differently from how the actually go throughout the argument: then (40) and (41) both become trivial, while the case for (39) is in no way weakened.

Our view preserves AGGLOMERATION: if each of a certain class of propositions is true at the closest accessible world where P is true, then at that world, all of those propositions are true. We think that the above argument against AGGLOMERATION fails because there is no single context in which all sentences of the form (39) are true. Each, nevertheless, is true in some contexts, and indeed has some tendency to evoke the kind of context in which it is true.

[. . .]

Chapter 2

Closeness

2.1 Three views

Here, again, is our bare-bones theory of conditionals:

CLOSEST A conditional with antecedent p and consequent q is true iff either there is no accessible p-world, or the closest accessible p-world is a q-world.

In saying this we are taking it for granted that where there are accessible p-worlds, there is a unique closest one. The aim of the present chapter is to say more about the relevant notion of closeness, and to motivate the claim that it obeys the required uniqueness assumption.

Some theories of conditionals use the ideology of worlds and closeness but make no assumption that there is always a unique closest accessible p-world when there is any accessible p-world. The most influential such theorist is Lewis (1973). Lewis's theory allows for two kinds of failures of uniqueness. First, there can be ties: there may be several maximally close accessible p-worlds (i.e. equally close p-worlds such that no p-worlds are closer than them). Second, there may be no maximally close accessible p-worlds: it could be that for every accessible p-world, there is a yet closer accessible p-world. Lewis's truth-conditions agree with ours in the case where there is a unique closest accessible p-world, but introduce a new element of universal quantification to deal with the other cases. When there are several maximally close accessible p-worlds, the conditional is true just in case all of them are q-worlds. When there are accessible p-worlds but no maximally close accessible p-worlds, the conditional is true just in case there is some accessible p-world such that every accessible p-world at least as close as it is a q-world. In fact the latter case also covers the case where there is

one or more maximally close p-worlds; so Lewis's theory can be stated as follows:

LEWIS A conditional with antecedent p and consequent q is true iff either there is no accessible p-world, or there is an accessible p-world such that every accessible p-world that is at least as close as that world is a q-world.[1]

Lewis himself only applied this schema to the analysis of counterfactual conditionals, but it is certainly worth exploring whether a theory with this shape could also work for indicatives.

Another important family of rival theories of conditionals makes use of the ideology of worlds and accessibility, but does not appeal to closeness at all in the specification of truth-conditions. On these 'strict conditional' accounts, the truth-conditions of conditionals are straightforward:

STRICT A conditional with antecedent p and consequent q is true iff every accessible p-world is a q-world.

(One might be tempted to think that at the schematic level we are presently working at, STRICT is perfectly compatible with our account: if we redefine 'accessible world' to mean what we previously meant by 'closest accessible world', won't STRICT then be just a terminological variant on CLOSEST? No: in the expression 'the closest accessible p-world', 'closest' is not playing the role of a monadic predicate of worlds. Being a closest accessible p-world is not just being (a) closest, (b) accessible, and (c) a p-world, just as the property of being a shortest spy is not the conjunction of the property of being shortest and the property of being a spy. Being a shortest spy is being a spy who is at least as short as any spy; being a closest accessible p-world is being a p-world that is at least as close as any accessible p-world.)

Lewis's argument against STRICT (for counterfactual conditionals) is well-known and prima facie compelling. Lewis notes that STRICT validates *Antecedent Strengthening*: whenever 'If P, Q' is true, 'If P and R, Q' is true (for any R). But this rule does not look to be valid. For example, the inference from

(1) If I bought my son a pet dog, he would be delighted

[1]Lewis avoids the need to say 'accessible' all the time by setting things up in such a way any accessible world is closer than all inaccessible worlds (if there are any inaccessible worlds). But this is not essential to the aspect of Lewis's theory that we are presently concerned with.

to

(2) If I bought my son a pet dog and strangled it, he would be delighted

seems invalid. Or what comes to much the same thing: the conjunction of the first with the denial of the second seems perfectly consistent and felicitous. We note that exactly the same kind of argument can be given for indicative conditionals: for example, the inference from (3) to (4) seems just as terrible:

(3) If he bought his son a pet dog, his son was delighted

(4) If he bought his son a pet dog and strangled it, his son was delighted.

To keep STRICT going in the face of these prima facie compelling considerations against it, its proponents have posited widespread context-shift as regards what counts as accessible. Their thought (*cite von Fintel*) is that in the context in which (1) is uttered truly, worlds where the speaker buys a pet dog and strangles it are inaccessible; but because of the presupposition of nonvacuity, uttering (2) triggers a new context in which at least one such world is accessible. The central mechanism that triggers contextual shifts is thus held to be that of presupposition accommodation. On this view, *Antecedent Strengthening* is valid in the sense that it preserves truth when context is held fixed; but realistic cases where the putative counterexample sequences are uttered are not cases where context is held fixed.[2] Since the views about accessibility we developed in chapter 1 also involve quite an extensive amount of context-shift, including context-shift driven by presupposition accommodation, we are in no position to discount STRICT simply on account of the contextualism it requires.

One feature that distinguishes CLOSEST from both STRICT and LEWIS is that status of the principle of Conditional Excluded Middle:

CEM Either if P, Q or if P, not-Q.

[2]Note that if one allowed context-shift to run completely rampant, one could even reconcile the truth-values delivered by STRICT with those delivered by CLOSEST. The idea would be that at the context where a conditional is uttered, the set of accessible worlds (in the sense relevant to STRICT) are all and only those accessible worlds (in the sense relevant to CLOSEST) that are at least as close as every accessible world where the antecedent is true. Unlike a version of STRICT that tries to try its account of context-shift to the familiar phenomenon of presupposition accommodation, this kind of mad-dog contextualism seems to depart so much from the standard way of thinking about what it means for distinct utterances to belong to the same context that evaluating it would require getting a lot clearer about what theoretical role is being assigned to the new conception of 'context'.

Given CLOSEST, instances of this schema will be true independent of context.[3] By contrast, STRICT obviously allows for interpretations of instances of CEM on which they fail to be true, because the set of accessible worlds includes both *P*-and-*Q* worlds and *P*-and-not-*Q* worlds. Likewise, LEWIS allows for two kinds of failures of instances of CEM. In one kind of case, there are several *P*-worlds tied for maximal closeness, which differ with regard to the truth-value of *Q*; in the other kind of case, there are no maximally close *P*-worlds, and for every *P*-and-*Q* world there is a closer *P*-and-not-*Q* world, and for every *P*-and-not-*Q* world there is a closer *P*-and-*Q* world.

Proponents of STRICT and LEWIS are well aware of this feature of their views; indeed their views have been partly motivated by the desire for CEM to come out invalid. However, we think there are strong reasons to like CEM. We will consider several such reasons in the present section; in section 2 we will address some arguments against CEM.

2.2 Chance and confidence-theoretic arguments for CEM

The most central considerations for us turn on facts about the chances of conditionals, and the levels of confidence we should have in conditionals. We are going to look at some instances of CEM involving fair coins, which seem like good test cases: if there were a problem for CEM, one would expect it to show up for conditionals concerning the outcomes of fair coin tosses.

Let's begin with chance (understood as objective rather than epistemic). Consider one of the coins currently in your pocket. What's the chance that it would land Heads if it were tossed in the next minute? Around 50%, surely (unless you are in the habit of carrying around trick coins). Similarly, the chance that it would fail to land Heads if it were tossed is around 50%. But chance is a kind of probability, and it a basic theorem of the probability calculus that the sum of the probabilities of two propositions equals the sum of the probabilities of their disjunction and their conjunction. And in this case, the conjunction of the two conditionals—namely that if the coin were tossed it would land Heads, and if the coin were tossed it would fail to land Heads—is absurd, and deserves zero credence. So the chance of the disjunction of the conditionals—which is an instance of CEM—is roughly one. In fact, it would seem to be *exactly* one, since any surprising factors

[3] At least if we continue take it for granted that every world is either a *Q*-world or a not-*Q* world. Chapter 4 will consider whether we sometimes need to invoke "incomplete" worlds that do not conformt to this generalisation.

that might elevate the chance of one disjunct above 50% would presumably reduce the chance of the other disjunct below 50% by the same amount. But given that the disjunction has chance one, it would be preposterous to deny that it is true.[4]

Another way to motivate CEM for counterfactuals is to consider the levels of confidence that seem to be appropriate. For example, concerning a certain untossed coin which you have no special reason to suspect of being a trick coin, it seems that you should be about 50% confident that it would have landed Heads if you had tossed it, and about 50% confident that it would have failed to land Heads if you had tossed it. After all, discovering that the coin and the table it is being tossed on are magnetised in such a way as to favour Heads looks like it should make you more confident that the coin would have landed Heads if you had tossed it. Discovering a setup of magnets that favours Tails would make you less confident. Without evidence for any such setups, it seems obvious that your credence should be middling. And for exactly the same reasons, it looks like you should be about 50% confident that the coin would have failed to land Heads if you had tossed it. Since you should be certain that the conjunction of these counterfactuals is false, and since rational levels of credence conform to the probability calculus, your level of confidence in the disjunction—an instance of CEM—had better be about 1. Indeed, it seems that it should be exactly one, since any reasons for raising your credence in one of the disjuncts above 50% seems like an equally good reason for lowering your credence in the other disjunct by the same amount.[5]

These ways of assigning credences to counterfactuals can be further supported by considering the motivating role of counterfactuals in deliberation. Suppose that you face an uncomfortable choice between opening two boxes. You are 49% confident that Box A contains a bomb primed to explode on opening the box, and 51% confident that it contains nothing. You know that Box B contains a bomb linked up to a fair coin: if the box is opened, the coin will be tossed and the bomb will explode if it lands Heads. Obviously you

[4] We wouldn't quite want to assume that *all* chance-one propositions are true: certain examples involving infinitely fine-grained outcomes make trouble for that simple generalisation. Similar points apply to arguments for the truth of a proposition from the premise that its epistemic chance is 1, or for the premise that we ought to assign it a credence of 1. But it seems hopeless to try to leverage these considerations into a defence of the denial of CEM.

[5] A closely related argument for CEM appeals to judgments about epistemic probability rather than about appropriate levels of confidence, e.g. that the epistemic probability that the coin would have landed Heads if it were tossed is about 50%.

should open Box A here. And the obvious explanation of this is that you are more confident that you would be killed if you opened Box B than that you would be killed if you opened Box A. But your credence that you would be killed if you opened Box A equals your credence that there is a bomb in Box A, namely 49%. So your credence that you would be killed if you opened Box B, which is equal to your credence that the coin in Box B would land Heads if it were tossed, looks to be over 49%–in fact, 50%. Standard CEM-deniers, by contrast, will think that insofar as you are confident that you won't take Box B, and hence confident that the counterfactual 'You would be killed if you opened Box B' has a false antecedent, your credence in that counterfactual should be very low. They will need some other, more complicated and (we think) less natural way of relating credences in counterfactuals to good deliberation.

These natural confidence-theoretic judgments are far out of line with the sorts of credences that seem to be recommended by standard versions of LEWIS and STRICT. For Lewis, cases involving chancy processes like coin-tossings are a primary motivation for allowing ties in the closeness ordering; given his actual theory of closeness, when we know that a coin is fair and was not tossed, we can be quite confident that the set of maximally close tossing worlds contain a mix of Heads and Tails worlds, and hence quite confident that each conditional are false. (Even if we aren't sure whether the coin was tossed or not, there will be some significant portion of our probability space devoted to the hypothesis that it was not tossed and that both conditionals are false because of ties.) Similarly, extant versions of STRICT tend to say things about accessibility that encourage the idea that the accessible worlds in this case include both Heads-landing and Tails-landing worlds, in which case both conditionals will come out false according to STRICT. We will certainly get this result if we interpret STRICT using anything like the conception of accessibility described in chapter 1. One our claims in that chapter is that the same notion of accessibility relevant to counterfactuals can also be picked up by certain modals like 'has to' and 'might have'; and obviously, given that the coin might have been tossed, it might have landed Heads, and might have landed Tails, and didn't have to land Heads, and didn't have to land Tails.[6]

[6]Of course, proponents of STRICT might attempt to preserve our motivating confidence judgments by introducing some devious new conception of accessibility under which, in the case of an untossed fair coin, the accessible worlds where the coin is tossed are all alike as regards how it lands, and our 50% confidence reflects our uncertainty about which worlds are accessible. We will briefly consider this view later. ***One point to note: given the kind of contextual shiftiness that such a view would require, we can't be thinking that accessibility is 'sticky' in the way required by the von Fintel/Gillies account of Reverse Sobel Sequences. We also can't assimilate accessibility restriction to the broad model of quantifier domain restriction in general, where the stickiness phenomenon is clearly a real thing.

We find a similar pattern of confidence-theoretic judgments when we turn from counterfactuals to indicative conditionals. Here the relevant confidence-theoretic judgments are ones we already discussed back in section 1.5, and their force has been widely acknowledged. For example, if you are not sure whether a certain coin was tossed yesterday, and have no special evidence favouring the hypothesis that it was tossed and landed Heads over the hypothesis that it was tossed and landed Tails, it seems that you should be about 50% confident that it landed Heads if it was tossed and about 50% confident that it landed Tails if was tossed, and hence—since you should be about 0% confident that it both landed Heads and landed Tails if it was tossed—you should be about 100% confident that either it landed Heads if it was tossed, or it landed Tails if it was tossed. So the confidence-theoretic case for CEM looks as strong for indicatives as for counterfactuals.[7]

While many opponents of CEM seem by our lights to have let their theory ride roughshod over their natural confidence-assignments, some of them have shown enough awareness of the ordinary practice to want to explain away data of the sort we have been relying on. Here is Jonathan Bennett:

> Admittedly, we often find it natural to say things like 'There's only a small chance that if he had entered the lottery he would have won', and 'It's 50% likely that if he had tossed the coin it would have come down heads'. In remarks like these, the speaker means something of the form $A > (P(C) = n)$—if the antecedent were true, the consequent would have a certain probability; yet the sentence he utters means something of the form $P(A > C) = n$.... When we use one to mean the other, we employ a usage that is idiomatic but not strictly correct. (Bennett p. 251.)

Bennett is not very clear about whether the relevant notion of chance here

[7]We haven't included an argument based on the objective chances of indicative conditionals, because it's not so easy to find sentences where it's clear both the relevant notion of chance is objective and that the conditionals are genuine indicatives—recall that we have suspended judgment on the status of 'does-will' conditionals. One could however try to make something of examples like 'There's a fifty-fifty chance that if he tosses this coin during the next hour, he wins a prize'.

is epistemic or objective. **check** But whichever way we go, we don't think the diagnosis is very promising. First of all, Bennett seems to be accusing us of conflating things that we seem in fact to quite good at distinguishing. This is particularly clear if we replace likelihood claims with claims about particular people's degrees of confidence, as in

(5) I am pretty confident that if this dice had been rolled without my knowing that it was rolled, it would have landed on some number other than 6.

The analogue of Bennett's move in this case would be to say that (5) is conflated with (6):

(6) If the die had been rolled without my knowing that it was rolled, I would have been pretty confident that it had landed on some number other than 6.

But it is hard to believe that we could conflate such obviously different claims. Moreover, even if such a conflation were plausible, there would be no prospect of using it to explain away our temptation to regard (5) as true, since (6) is manifestly false.

Even confining our attention to claims of chance, as Bennett does, the "conflation" strategy strategy delivers terrible results. The problem is especially clear when the time-index of the chance ascription is made explicit:

(7) It is likely right now that if Jim drank arsenic tomorrow, he would be dead by the weekend.

Bennett's transformation would turn this into

(8) If Jim drank arsenic tomorrow, it would be likely right now that he would be dead by the weekend.

Assuming that it is not in fact likely right now that Jim will be dead by the weekend, (8) seems false: it's not true that if he drank arsenic tomorrow, he would have all along been likely to have been dead by the weekend. And this is so whether the relevant notion of likelihood is understood as objective or epistemic. In neither case are the probabilities today plausibly regarded as counterfactually dependent on how things play out tomorrow.

[. . .]

2.3 CEM and denying conditionals

Another group of considerations in favour of CEM have to do with the interaction of negation and denials with conditionals. The validity of CEM is equivalent to that of the inference from 'It is not the case that if P, Q' to 'If P, it is not the case that Q'. However such inferences are hard to evaluate directly: after all, explicit negations of the form 'It is not the case that if P, Q' and 'It is false that if P, Q' are not common in natural language, and so it would be unwise to place too much weight on any instincts regarding sentences of that form. But there are forms of denial that are much more natural. For example, we can consider negative answers to questions concerning a conditional:

(9) Would this coin have landed Heads if it had been tossed?

— No.

This answer is clearly quite inappropriate unless you have some very unusual knowledge about the characteristics of the coin or the tossing setup. Furthermore, in any setting where that answer is felicitous, one is also in a position to assert

(10) If the coin had been tossed, it would have landed Tails

assuming that one's insider information does not disrupt the standard assumption that if the coin had been tossed, it would have either landed Heads or Tails and not both. But the standard versions of STRICT and LEWIS just described seem to entail that merely knowing that the coin was fair would be enough to justify the negative answer to (9), and that the inference from this negative answer to (10) is completely unjustified.

This kind of consideration also supports CEM in the case of indicatives. For example, the reasoning of the client in the following dialogue looks cogent:

(11) *Client:* Will I have a big tax bill if I have won the lottery?

Accountant: No.

Client: Great: so if I have won the lottery my financial troubles are completely over and done with.

In saying 'No', the accountant is clearly committed to 'You won't have a big tax bill if you have won the lottery'.

It might be suggested that the inappropriateness of the negative answer to (9) should be assimilated to the phenomenon of 'Neg-raising' that applies to words like 'believe' and 'want', wherein sentences in which negation takes wide scope over some other operator fail to entail, but seem in some other way to suggest, the truth of the corresponding sentences in which negation takes narrow scope. For example, saying 'I don't believe it is raining outside', or answering 'No' to the question 'Do you believe that it is raining outside?', tends to suggest that you believe it isn't raining outside. Likewise 'I don't want to go to London' suggests 'I want not to go to London'. Could this mechanism be enough to explain the seeming goodness of the inference to (10) from the negative answer to (9), and the infelicity of that negative answer? We doubt it. The suggestions associated with Neg-raising are easily cancelled: for example, if my answer to the question 'Do you want to go to London?' is 'No, I don't care either way', no-one would be tempted to infer that I want not to go to London. By contrast, there is nothing which one could to 'No' in answering (9) which would make it felicitous and block the inference to (10). For example, saying 'No, it's a fair coin' does nothing to make the answer any better.

Some have argued that when we put focal stress on the word 'would' we get judgments more in line with those of CEM-deniers. Hajek (2007), for example, claims that (12) sounds true:

(12) It is not the case that the coin WOULD land tails if it were tossed.

We have no clear judgment about what is going on in this sentence; we suspect that sentences generated by prefixing conditionals with 'It is not the case that' are so unnatural that our views about them are especially likely to be theory driven. One might try to avoid this by using the question-answer format. But it doesn't seem that focusing 'would' in (9) makes the answer 'No' much more acceptable. (The most obvious reason for focusing this 'would' is to signal one's challenge to a previous assertion of 'The coin would have landed heads if it had been tossed': this context certainly does nothing to improve the negative answer.) Perhaps, however, we can find some more natural examples where focusing 'would' provides evidence against CEM, at least for whatever reading of the conditional is activated by such focus. Suppose for example that have a pile of coins, some normal, some double-headed and some double-tailed. It is not unnatural to say things like (13) in describing this situation

(13) There are three kinds of coins in this pile: those that WOULD land heads if tossed, those that WOULDN'T land heads if tossed, and the rest.

And once you have got into this mood, you can for example, issue the instruction

(14) Just give me the coins that WOULD land heads if tossed

where one expects obedience to consist in handing over the double-headed coins.

One point to make about these focus-based examples is that they carry across to 'will', including uses of 'will' that are not embedded in conditionals. Compare (13) and (14) with:

(15) There are three kinds of students: there are the ones that WILL pass, there are the one's that WON'T pass, and there are the students that might go either way.

(16) Don't give a student a lot of time unless they WILL pass

This provides a reason for caution about the putative anti-CEM data from focus, since even those philosophers who reject CEM for counterfactuals will likely still want to preserve sentences like 'Either you will pass this exam or you will not pass this exam'. Granted, there is the radical option of adopting a theory of 'will' according to which it involves universal quantification over a range of possible futures, and hence fails to commute with negation even when only one future time is relevant. But we suspect that the kind of focus-driven effect on display is something much more general, that does nothing special to illuminate the semantics of 'would' or 'will'. For example, we are not seeing a big difference between the foregoing examples and the following:

(17) There are three kinds of patients in this ward: those that ARE sick, those that AREN'T sick, and those that might or might not be sick.

(18) Don't waste any more time doing tests on the patients who ARE sick.

A tentative hypothesis: putting focal stress on bland small words like 'are' and 'would' can make the relevant clause behave as if it were prefixed by some epistemic operator like 'It is known that. . . ' or 'It is knowable that. . . '.

2.4 Other arguments for CEM

[. . .]

Some authors who reject the claim that instances of CEM are invariably true have suggested that they nevertheless have a different positive status, namely that of being *true whenever their presuppositions are satisfied*. The idea is that 'If P, Q' presupposes what the corresponding instance of CEM, namely 'Either if P, Q or if P, not Q', expresses. Most proponents of this idea have been advocates of STRICT, so that for them, the content of this presupposition is that all accessible P-worlds are alike with respect to whether or not they are Q-worlds: the presupposition is thus in their hands a "presupposition of homogeneity". A favoured analogy is with plural definites it's suggested that 'The philosophers left the room' presupposes that either all or none of the (relevant) philosophers left the room, so that 'Either the philosophers left the room or the philosophers failed to leave the room' is true whenever its presuppositions are satisfied.[8]This move gives instances of CEM the status of "Strawson validity": true on any uniform interpretation on which everything they presuppose is true.

As noted in section 1.4, Strawson-validity confers nothing like the same kind of intuitive security as validity proper. For example, the following sentences are Strawson valid given standard tenets about the presuppositions of 'the' and 'stopped':

(19) a. The elephant wearing a hat in my bedroom is wearing a hat.

b. Either John has stopped shouting, or John is shouting right now.

c. Either I regret eating the moon, or I ate the moon and do not regret it.

It is thus important not to assimilate the kind of thorough embrace of CEM that we have been advocating to the more lukewarm endorsement that goes along with Strawson-validity.[9] Nevertheless, we agree that there are some

[8]Citations: von Fintel 1999; 2001, Klinedinst, Kriz.

[9]The contrast between Strawson-validity and the stronger form of validity tends to be obscured in theories according to which presupposition failure is held to make for truth value gaps. If one holds this, then not even the law of non-contradiction is sacrosanct, since it will have instances that fail to be true because of presupposition-failure, such as 'It is not the case that (John regrets eating the moon and it is not the case that John regrets eating the moon)'. Within this framework, then, one might think that Strawson-validity was the strongest status that could reasonably be claimed for any schema. However, even within this framework, one can still draw a contrast between schemas which are guaranteed to be true whenever non-gappy expressions are substituted for the schematic letters and other schemas whose Strawson-validity is due to the presupposition-theoretic properties of the non-schematic expressions in the schema.

cases where a diagnosis of Strawson-validity provides the best all-things-considered explanation of the positive felt status of some schema. Indeed, we have provided such a diagnosis in the case of the following good-looking schema:

CNC Not ((if P, Q) and (if P, not Q))

CNC* Not (if P, Q and not Q)

On our account, these is false when there are no accessible P-worlds; however, in such cases all the ingredient conditionals—and hence also their conjunctions and negations—will suffer from presupposition failure thanks to the presupposition of non-vacuity.

The thesis that CEM is Strawson-valid gives proponents of STRICT or LEWIS a story about the goodness of some of the inferences that feature in certain of our subsidiary arguments against those views. For example, the fact that answering 'No' to the question 'Would this coin have landed Tails if it had been tossed?' is inappropriate if one takes the coin to be fair (and not to have been tossed) could be accounted for by saying that such an answer would signal aquiescence to the presupposition of the question (that either the coin would have landed Tails if it had been tossed or the coin would have failed to land tails if it had been tossed), a presupposition which is false if the coin is fair and untossed. However, the Strawson-validity of CEM does not even begin to make sense of the facts about chances and credences that drive our central argument against STRICT and LEWIS. For example, on those approaches, 'I am pretty confident that if he had rolled the die it would have come up between 1 and 5', as uttered in a setting where one is pretty confident that the die is fair, will be like 'I am pretty confident that the elephant in the room is more than five feet tall', as uttered in a setting where one is pretty confident that there is no elephant in the room.

We can also probe more directly the question whether conditionals carry the presuppositions that would be required for CEM to be Strawson-valid given STRICT or LEWIS. In general, when a sentence P carries a presupposition, 'S doesn't know whether P' carries the same presupposition: for example, 'Bert don't know whether Alice stopped smoking' presupposes that Alice

used to smoke.[10] Given this presuppositional profile, then, the Strawson-validity approach ought to predict that that claims of the form 'S doesn't know whether the coin would have landed Heads if tossed' would be infelicitous when the coin in question is known to be fair, or indeed even when it is not known not to be fair. In fact, however, such claims seem completely fine under such circumstances. In our view, such attributions of ignorance are acceptable even in the case of a coin that's known to be fair; but their felicity may be even more evident when fairness is only of the epistemic possibilities. Suppose for example that Sally picked her coin from the bucket with a mix of fair and double-Headed coins—'Sally doesn't know whether her coin would have landed Heads if tossed' seems like an excellent description of this situation. Or consider a self-attribution of ignorance like 'I don't know whether he would have said yes if she had asked'. Here it is fine to follow up with 'He is so unpredictable: he might have said yes and he might not have'. There is a contrast here between conditionals and the case of the plural definites that forms the inspiration for the view: 'I don't know whether the philosophers left the room' does still seem to carry some suggestion that the philosophers moved or stayed as a bloc. Note also that these tests for presuppositionality give a much more favourable verdict in the case of CNC and CNC*: 'I don't know whether cannons would both work and not work if Aristotelian physics were true' is quite bizarre.[11]

2.5 Closeness and similarity

In the literature—especially thanks to the influence of Lewis (1973)—closeness is standardly understood as some kind of similarity. For one world to be closer than another is for the former to be more similar to the actual world than the latter in the relevant respects. What the relevant respects are and what weight they carry is supposed to be up for grabs: Lewis emphasises that the respects and the weighting might not be recoverable simply from general pre-theoretic ideas about overall similarity between worlds. Lewis also thinks that there is plenty of context-sensitivity as regards respects and weightings. Nevertheless, the use of the word 'similar' isn't meant to be

[10] It also seems to presuppose that Bert knows that Alice used to smoke, though that won't be important here.

[11] And following up this speech with 'That's because Aristotelian physics is impossible' would be completely bizarre, although the presuppositional view of CEM suggests that this should be the analogue of the 'He is so unpredictable' followup mentioned above.

utterly divorced from its home in ordinary language. There are various structural expectations which are triggered by the ideology of similarity which are preserved on Lewis's view: the kind of relation we are supposed to be thinking about is one specifiable by a scale that comes in degrees, where the degrees in question are determined by some aggregation procedure whose inputs are degrees of resemblance in a range of specified respects, where resemblance comparisons in the particular respects are supposed to be much more straightforward.[12] And even this very schematic notion of similarity would lead one to expect there to be many cases where two worlds are equally similar to a third world. After all, the relations of resemblance in particular respects that enter into the final aggregation procedure are generally things that do allow for ties: for example, two worlds could be equally similar to a third in respect of *mass* even when one was greater than it in mass and the other less. Moreover, if the aggregation procedure ever allows tradeoffs between different respects of resemblance without one trumping the other, one would expect there to be cases where the tradeoff results in an exact balance. In view of the richness of the space of possible worlds, it is thus hard to see how anything we could think of as a relation of overall resemblance could fail to generate many ties.

If the closeness relation is understood as one of similarity, there is also considerable pressure to allow for failures of the Limit Assumption. Certainly for particular respects of resemblance where comparisons are straight-

[12] As Goodman (???) points out, some of the structural expectations invoked by Lewis's use of the word 'similar' are actually ones which he disavows in his most careful moments in a way that many expositors have overlooked, including Lewis himself in some of his less careful moments. In particular, if one pronounces the semantically relevant three-place relation as 'w_2 is more similar to w_1 than w_3 is', one would expect that the following pattern can never arise:

- w_2 is more similar to w_1 than w_3 is
- w_3 is more similar to w_2 than w_1 is
- w_1 is more similar to w_3 than w_2 is

For, letting $|ww'|$ be the degree of similarity between the worlds w and w', these claims would seem respectively to entail the jointly unsatisfiable $|w_1w_2| < |w_1w_3|$, $|w_2w_3| < |w_1w_2|$, and $|w_1w_3| < |w_2w_3|$. But in fact, Lewis is open to the idea that the relevant three-place relation does contain triples with this cyclical structure, and indeed, as Goodman argues, such cases can arguably be constructed for the particular similarity relation described in Lewis 1979. The key to resolving the mystery is that for Lewis, the key three-place similarity-theoretic relation should really be pronounced something like 'w_2 is more similar to w_1 than w_3 is in the respects that matter at w_1'; differences in which respects "matter" at different worlds can then create the surprising cycles.

forward, there can be propositions such that for any world where that proposition is true, there is another world where that proposition is true that resembles that actual world more in the relevant respect. Take total mass as the relevant respect, and consider the set of worlds whose mass is greater than m_1, where m_1 is greater than the mass of the actual world. Clearly for any world in this set, there is another world that resembles actuality more in respect of total mass, e.g. one whose total mass is halfway between m_1 and that world's mass. And given standard assumptions about the continuity of various fundamental magnitudes it is hard to see how this sort of phenomenon could fail to be replicated at the level of overall similarity. Thus, even prior to its details being filled in, a similarity-driven conception of closeness strongly suggests both kinds of Lewisian counterexamples to CEM.

Our response is to completely jettison the similarity-driven conception of closeness. We have various reasons for going this route. Two of them are broadly logical. First, as discussed in section 1.6, infinite agglomeration is just as compelling as finite agglomeration, but as we have just seen, identifying the closeness ordering with a similarity ordering makes the Limit Assumption and thus infinite agglomeration look untenable. Second, as discussed in the previous section, we think there are good reasons to accept CEM, which is again threatened by identifying the closeness ordering with a similarity ordering.

Even if one accepted these logical principles, one might think it was an overreaction to abandon wholesale the connection between similarity and closeness: one might might propose a picture where similarity facts constrain but do not determine closeness facts, or more generally, a picture where there is a general tendency for the similarity and closeness orderings to line up. (If we only had to deal with the problem of ties, we could entertain the obvious constraint that whenever one world is is more similar to actuality than another it is closer, treating the closeness ordering as a mere tie-breaking refinement of the similarity ordering—cf. R. C. Stalnaker 1981. It is far less clear how we could understand the constraining role of similarity if we think that the closeness ordering does, while the similarity ordering does not, respect the Limit Assumption.)

But what really pushes us to reject even this moderate attitude towards the relevance of similarity to closeness is the the pattern of confidence judgments we find for both counterfactuals and indicatives. As a warmup, let us begin with a familiar kind of concern. Suppose first that we tossed a fair coin ten times yesterday and saw that it landed Heads on four of those times. Consider the counterfactual

(20) If the coin had landed Heads nine times, it would have landed Heads all ten times.

Our judgment is that the right way to assign confidence to the proposition expressed by (20) is to consult what you know about the conditional chances: of the 2^{10} equiprobable ways the coins could have landed, ten involve exactly nine Heads outcomes while one involves ten Heads outcomes, so your credence in the proposition should be 1/11. But if the kind of similarity that constrains closeness is understood in a pre-theoretically natural way, one would expect that any world where the coin comes up Heads ten times will be less similar to actuality than at least one world where it comes up Heads only nine times, so that (20) would get a vanishingly low credence. For essentially similar reasons there will be a mismatch in the case where you don't know how the coin landed. So long as the coin is fair we think the appropriate credence is still 1/11; but the similarity-constrained approach threatens to entail that the only way (20) could be true would be for ten Heads to in fact have been tossed, a hypothesis to which we assign a much lower credence (namely $1/2^{10}$). (Unless the coin actually landed Heads every time, then for every ten-Heads world we should be able to find a nine-Heads world that is more similar to actuality.)

The general problem is that once we connect closeness to similarity—even in the more modest, constraining way—we will be left with odd confidence distributions. In particular, when there are various ways of a thing happening some of which are more similar to actuality than others, the similarity will push us to be unreasonably confident that if that thing had happened it would have happened in one of the more similar ways. Essentially this objection was made in an early review of Lewis's *Counterfactuals* by Fine (???): Fine considered the counterfactual 'If Nixon had pressed the button there would have been a nuclear holocaust', and objected that Lewis's theory yields to the problematic judgment that this is false, since worlds where the button is pressed and some subsequent misfire blocks the expected global nuclear war are more similar overall to the actual, nuclear-war-free world. In response, Lewis makes the point that the relevant similarity metric does not have to coincide with the one elicited by pre-theoretic similarity judgements about worlds, and denies that the relevant peaceful worlds are any more similar to actuality than the war worlds on his *intended* similarity

metric.[13] But structurally the same problem recurs if we look at the respects of similarity that really do matter according to Lewis. For Lewis, the most important consideration (besides the avoidance of large, widespread exceptions to the actual laws of nature) is the extent of the spatiotemporal region of perfect match. When the region of perfect match between w_3 and w_1 is a proper part of the region of perfect match between w_2 and w_1, and neither w_2 nor w_3 contains 'big miracles' with respect to the laws of w_1, w_2 is more similar to w_1 than w_3 is. If closeness is constrained by a similarity ranking that works in this way, the upshot is that we can be confident, generally speaking, that if things had gone otherwise than they actually go, the departure would have been later rather than earlier (since the later the departure, the larger the initial chunk of perfectly matching spacetime). But this still leads to problems in the coin example. Suppose that the coin landed Tails the first time it was tossed. Then on Lewis's account some worlds where it lands Heads exactly nine times—namely, those which match actuality throughout a period including the first toss—are more similar to actuality than any world where it lands Heads all ten times. So if we know that it landed Tails the first time, we should be practically certain that (20) is false. Moreover, if we don't know anything about how it landed but know that it was fair, given that we should still be at least 50% confident that it landed Tails the first time, the only way we could generate the intuitively correct credence of 1/11 in (20) would be to assign (20) a credence of 2/11 conditional on the hypothesis that the coin landed Heads the first time, which seems completely crazy. (After

[13] Lewis's final similarity ranking is intended not merely to prevent the peace worlds from counting as more similar to actuality than the war worlds, but further to get some of war worlds to be more similar than any of the peace worlds. We have further concerns about whether his official specification actually achieves this goal. Lewis's trick for promoting the war-worlds above peace worlds—assuming determinism—is to say that all of the latter contain at least two small 'miracles', i.e. localised exceptions to the actual laws, whereas some of the former contain only one small miracle. We are not sure why Lewis is so confident that a delicately adjusted small miracle couldn't do the job of simultaneously ensuring a button-pushing and its failure. For example, Nixon could trip and fall in such a way that his finger hits the button just after his teeth sever the wire connecting it to the nuclear arsenal. (The problem is even more obvious in a case where there are various ways of pressing the button, a few of which are ineffective.) One could however tinker further to address this worry, e.g. by saying that the aforementioned tripping-and-falling small miracle is more "remarkable" than certain miracles that lead to war, and for this reason makes for more dissimilarity to the actual world. (This suggestion would be in the spirit of some of Lewis's own ideas about the generalisation of the account to the case of indeterminism which we will discuss below.) By contrast, the problem for Lewis's account which we focus on in the main text is not one that could plausibly be evaded by small adjustments to the similarity ranking.

all, conditional on the hypothesis that the coin landed Heads the first time, we are certain that (20) is true just in case the coin would have landed Heads on all of the final nine flips if it had landed Heads on at least eight of those flips. Intuitively, we ought to assign the latter proposition a credence of 1/10, both unconditionally and conditional on the hypothesis that the coin landed Heads the first time.)

The problem here is essentially the same as the problem of Pollock's Coat which we discussed in chapter 1. There, we saw that a similarity metric that favours late divergence will license us to be very confident in the truth of

(21) If my coat had been stolen last year it would have been stolen on December 31st.

when we know that our coat was not in fact stolen last year. And for reasons similar to those discussed above, an approach that licenses such confidence will also generate unreasonable-looking credence profiles in propositions such as (21) in cases where we are uncertain whether our coat was stolen. In fact, we think that the appropriate way to assign credence to the proposition expressed by (21) on its most obvious interpretation is to set this credence equal to our expectation for the conditional chance of the coat being stolen on December 31st conditional on its being stolen some time during the year.[14]

Another respect of similarity that proponents of similarity constraints on closeness have actually regarded as important is similarity with respect of the *absence of remarkableness*, particularly in respect of the outcomes of chance processes. The task these theorists set themselves is to craft a notion of similarity on which, assuming that a certain plate is not in fact dropped, some worlds where it is dropped and hits the floor are more similar to the actual world than any world where it is dropped and quantum-tunnels right through the floor. On the (realistic) assumption that such events of quantum-tunnelling always have some miniscule but nonzero chance of happening, the idea that the disruption of *laws* is a count against closeness does not achieve the desired result, since no actual laws need be disrupted by either kind of world. Clearly the worlds in question need not differ with respect to the size of the region of exact match, so Lewis's idea about the primary importance of exact match doesn't help either. Saying that *low-probability*

[14] Calculating this number is not straightforward given that the coat being stolen on one day requires it not to have been stolen earlier. In the case where you have no relevant discriminating evidence it will be less than 1/365 but certainly not zero.

events count against similarity is a non-starter: whatever happens, the precise trajectory of the plate will be a low-probability event, maybe even a zero-probability event. Lewis's suggestion is that among the various low-chance events we should distinguish a subclass of "remarkable" ones, and to count only these as making for dissimilarity. There are various ways of making this suggestion more precise. One possibility worth mentioning is to understand the remarkable outcomes within a given partition of possible outcomes to be those low-probability outcomes that are much more *natural* than most of the low-probability outcomes in the partition (using something like the notion of relative naturalness for properties developed in Lewis 1983). For example, the property of being a sequence of ten coin-tosses all of which land Heads looks considerably more natural than the property of being a sequence of ten coin-tosses which land in the pattern HTTHTHH-THH.[15] Similarly, consider any specification of a way for the plate to hit the floor and break that is detailed for the probability of the plate doing *that* to be roughly as low as the probability of it quantum-tunnelling: plausibly, each of the aforementioned specifications defines a property far less natural than the property of being a quantum-tunnelling through a potential barrier of such-and-such strength.[16]

Any attempt to specify a similarity relation that turns on considerations of remarkableness will have to reckon with two facts: (a) Assuming that the actual world is quite extensive, we can be very confident that many remarkable events actually occur; and (b) if things had gone differently in some specified way, it would still have been very likely for the ensuing history of the world to contain many remarkable events that don't actually occur. There is plenty to say here (see Hawthorne ???), but we think the best bet for the remarkableness lover is to add a dose of contextualism to the story. For example, we could allow context to contribute a "reference

[15]Of course both of these properties are a long way from being perfectly natural; but this need not disrupt the judgment of comparative naturalness.

[16]The naturalness-theoretic gloss on "remarkableness" brings it close to the notion of "atypicality" invoked by Williams (???), drawing on earlier work by Elga (???). The notion of atypicality invoked by these authors is rooted in a mathematical characterisation of the contrast between "random" and "non-random" infinite sequences of coin-tosses due to Gaifman and Snir (???). Williams hopes that Gaifman and Snir's insight can be generalised to finite sequences, and even to particular localised events, and uses expressions like 'simplicity' in characterising this generalisation. But Williams mainly focuses on sequences of coin-tosses; it is not straightforward to extract a particular treatment of the quantum-tunnelling case from his remarks.

property", such as the property of being a flip of a certain coin by a certain person during a certain period, or a pattern of motions of a certain plate during a certain period. The extension of this property gives us a reference class which varies from world to world, and may be sometimes empty. We could then articulate a remarkableness-driven respect of similarity such as the following: w_2 is more similar to w_1 than w_3 is iff the reference class at w_3 has more remarkable properties than the reference class at w_1 or the reference class at w_2.[17] (Of course a final story will have to say something about how this respect of similarity is to be aggregated with others.)

The structural problem we have identified for similarity-based constraints on closeness recurs for the remarkableness-based similarity measures. The basic problem, as always, is that similarity constraints mandate giving more credence to counterfactuals whose consequents characterise non-actual outcomes that are more similar to actuality, in cases where considerations of chance push in a different direction. Suppose for example that a fair coin landing heads or tails a hundred times in a row count as remarkable outcomes, while a certain specific sequence S of heads and tails outcomes counts as unremarkable. In the actual world we know that the coin was never tossed. Before the coin-tossing was called off, Fred bet that the coin was going to either land all Heads, all Tails, or in sequence S, and George bet that it was going to land in sequence S. Given a remarkableness-based similarity constraint, we should then be able to know that worlds where the coin is tossed a hundred times and the sequence of outcomes is S are closer than worlds with a hundred heads or with a hundred tails.[18] And we should thus be in a position to assign very high credence to the proposition expressed by (22):

(22) If Fred had won his bet, George would have won his bet too.

[17]Note that this toy theory counts w_2 as more similar to w_1 than w_3 in the relevant respect even if the reference class has ten remarkable properties at w_1, eleven at w_3, and only one at w_2. A more mechanical similarity measure based on counting remarkable properties would not give this result. But the failure to demote w_2 seems desirable. Suppose that in the actual world, I threw ten plates at one wall and they all quantum-tunneled. We don't want to be able to say 'If I had thrown the plates at the opposite wall, then at least one of them would have quantum-tunnelled'.

[18]Of course what matters according to similarity-lovers is overall similarity, not just the particular remarkableness-based respect of similarity that we are currently considering. But in this case, none of the other respects of similarity that have been thought to play a role—perfect match, approximate match, lack of miracles, typicality of the world as a whole, and so on—do anything to favour the all-heads or all-tails worlds over S-worlds, so the S-worlds will plausibly be counted as more similar overall as well as in respect of remarkableness.

But intuitively this proposition deserves a credence equal to the conditional chance of George's winning given Fred's winning, namely 1/3.[19]

The problem with similarity-constraints on closeness is thus quite general. But without a general theory of closeness that guarantees that the worlds where the plate hits the floor and breaks are closer than worlds where it quantum-tunnels through the floor, how are we to explain the assertability of (23a) and (23b)?

(23) a. If the plate had been dropped, it would have broken

b. If the plate had been dropped, it would not have gone right through the floor

The first point to note about these sentences is that, if we form degrees of confidence to counterfactuals in the manner endorsed in section 2.2, we should have a high degree of confidence in the propositions expressed by (23a) and (23b), assuming that we know that the antecedents are false and that the conditional objective chances of the antecedents on the consequents were high.[20] Of course, high credence isn't in general sufficient to explain assertability—consider 'This is a losing lottery ticket'. Perhaps assertability requires knowledge; in any case, the results of prefixing 'I know that' to (23a-b) are also assertible. How, without some general analysis of closeness in terms of similarity, or some analysis of accessibility that guarantees that there aren't any accessible worlds where the plate is dropped and tunnels through the floor, could we explain our ability to know that the closest accessible world where the plate is dropped isn't a tunnelling world? A good starting point for thinking about this question is the assertability, despite small objective chance of error, of simple claims about the future: for example, our background knowledge about quantum tunnelling does not block us from asserting 'This plate will soon be broken', or from self-ascribing knowledge that the plate will soon be broken. It is not an easy task to come up with a

[19]Could we avoid the undesirable result that (37) is true by appeal to some further contextualist trick, according to which (37) for some reason evokes a context in which the all-heads and all-tails outcomes don't count as remarkable in the way that detracts from their similarity? In principle yes, but it's hard to see how such an account would go. For example, saying that the mere *salience* of a particular low-probability outcome makes that outcome stop counting as remarkable will disrupt the proposed explanation of the truth of 'If the plate had been dropped it wouldn't have quantum-tunnelled through the floor'.

[20]Even if we don't know that the antecedent is false, the counterfactuals will still deserve high credence so long as we don't have evidence relevant to their consequents that is "inadmissible" with respect to the relevant time.

workable theory that pinpoints the relevant difference between future lack of quantum tunnelling and future lottery-losing. (For some ideas about this, including ideas according to which our reactions betray deep-seated errors, see Hawthorne 2004.) We can remain neutral about this here: the main suggestion we want to make is that one should approach the knowledge and assertability of counterfactuals in the same way as the knowledge and assertability of claims about the future. Whatever explains the unassertability of 'I will lose the lottery' will also explain the unassertability of 'If I had bought a ticket in that lottery I would have lost'; whatever explains why the unassertability of the first claim doesn't carry over to all sorts of other claims about the future which have small chances of being false will explain why the unassertability of the second claim doesn't carry over to all sorts of other counterfactuals whose consequents have a small chance of being false conditional on their antecedents.[21] Of course whatever defensive manoeuvres we make, we should concede that in the future case, if you are unlucky and the plate does in fact tunnel through the floor, then the earlier self-ascription of knowledge was false after all; similarly, if you are unlucky and the closest accessible world where the plate falls is one where it tunnels, your self-ascription of knowledge that the plate would break if it fell is false.[22]

The challenge of connecting up facts about the appropriate degrees of confidence in conditionals with facts about assertability and knowledge

[21]Here we are in agreement with Moss (forthcoming).

[22]In the case of future-tense claims, the prima facie attractive idea that knowledge is closed under many-premise deductions leads to further puzzles: by performing a conjunction introduction based on many intuitively knowable high-chance premises, one can come to know a conclusion that one knows to have very low chance. This is both odd in its own right, and makes further puzzles for the project of formulating an account of the rational connection between a proposition p and the proposition that p has a certain objective chance. On one approach, one's knowledge of p's low chance means that one *ought* to have a low degree of confidence in p (despite the fact that one in fact knows p and thus presumably *in fact* is quite confident in it). On a different approach, one ought to have a high degree of confidence in p (despite the fact that one knows its chance is low). In the terminology of Lewis's Principal Principle (Lewis 1980), the latter option will involve claiming that one's evidence at the time in question is "inadmissible" at that time. All of this structure carries over to the case of counterfactuals. Given multi-premise closure and (finite) agglomeration for counterfactuals, we will sometimes be in a position to know the proposition that if p were true q would be true despite also knowing that its current objective chance is low. Possible reactions include giving up multi-premise closure; tolerating the idea that one can know things to which one ought to have assigned low confidence; and tolerating the idea that we should sometimes be confident in something whose current chance we know to be know.

arises for indicatives as well as for counterfactuals. However, for indicatives, there is more scope for explaining assertability by appeal to the accessibility parameter. Suppose that we're in a context where it takes for a world to be accessible is that be compatible with my knowledge, and that I know that the plate won't be dropped without breaking. Then 'If the plate is going to be dropped, it is going to break' is guaranteed to be *true*. And if I *know* I know that the plate won't be dropped without breaking, I can know the proposition expressed in context by the conditional without having to draw at all on knowledge of closeness. On our view this kind of explanation is often apt, but is not the only path to knowledge of indicative conditionals: one can also draw on whatever resources one used to make room for knowledge of counterfactuals like (23a-b).[23]

2.6 Metaphysical worries

We have rejected similarity-based analyses of closeness. And as will be becoming increasingly clear to the reader, we are not going to put anything their place: we will be "treating the concept of closeness as primitive". At least, we will not be offering the kind of analysis of closeness that would allow us to break out of the circle of concepts that includes closeness as well as conditionals: on our account, so long as we are in a context where both w_1 and w_2 are accessible, 'w_1 is closer than w_2' is equivalent to 'If one of w_1 or w_2 were actualised, w_1 would be actualised' and to 'If one of w_1 or w_2 is actualised, w_1 is actualised'. We are a little wary of calling this an 'analysis', in part because we are a bit unclear about what it takes for something to count as an analysis. Claims of analysis are sometimes understood to carry implications about what grounds what, or what is more metaphysically fundamental or natural than what, or what is more conceptually basic or "explanatorily prior" to what: we are not endorsing any such claim or priority for closeness over conditionals.

So, in a sense we are less ambitious in our goals than many other theorists of conditionals have been. Nevertheless, the claims we are making are by no means trivial - indeed as we have seen they are inconsistent with the views of many other writers, so we feel no need to defend the substantiveness or interest of the views we are putting forward.

"Treating closeness as primitive" is merely refraining from offering any

[23] By contrast, it is much harder to see how a proponent of STRICT can make room for knowledge of indicative conditionals in cases where iterated knowledge is unavailable.

definitions or equivalences that break out of the circle just noted. It is not any kind of claim about closeness. We are thus not making any claim to the effect that closeness is fundamental or perfectly natural or unanalysable or anything else of the sort. Nevertheless, the structural claims we are making about closeness invite a certain hard-to-pin-down metaphysical worry. Take, for example, the claim that a certain untossed coin is either such that some world where it lands Heads is closer than any world where it lands Tails, or such that some world where it lands Tails is closer than any world where it lands Heads. What, people want to know, could conceivably *make it be the case* that one rather than the other of these two possibilities obtains? In what could the difference between the untossed toins that would have landed Heads if tossed and the ones that would have landed Tails if tossed conceivably *consist*?

[. . .]

Bibliography

Anderson, Alan (1951), 'A Note on Subjunctive and Counterfactual Conditionals', *Analysis*, 12: 35–8 [11].

Condoravdi, Cleo (2002), 'Temporal Interpretation of Modals: Modals for the Present and for the Past', in *The Construction of Meaning*, ed. David Beaver, Stefan Kaufmann, Brady Clark, and Luis Casillas, Stanford: CSLI Press, 59–87 [18].

Dorr, Cian (2016), 'Against Counterfactual Miracles', *Philosophical Review*, 125: 241–86 [9].

Fine, Kit (2011), 'Counterfactuals Without Possible Worlds', *Journal of Philosophy*, Forthcoming in *Journal of Philosophy* [50].

Gibbard, Allan (1981), 'Two Recent Theories of Conditionals', in *Ifs: Conditionals, Belief, Decision, Chance, and Time*, ed. William L. Harper, Robert Stalnaker, and Glenn Pearce, Dordrecht: D. Reidel Publishing Company [8].

Harper, William L., Stalnaker, Robert, and Pearce, Glenn (1981) (eds.), *Ifs: Conditionals, Belief, Decision, Chance, and Time* (Dordrecht: D. Reidel Publishing Company).

Hawthorne, John (2004), *Knowledge and Lotteries*, Oxford: Oxford University Press [75].

Heim, Irene (1991), 'Artikel und Definitheit', in *Handbuch der Semantik*, ed. Arnim von Stechow and D. Wunderlich, Berlin: De Gruyter [11].

Heim, Irene and Kratzer, Angelika (1998), *Semantics in Generative Grammar*, Malden, MA: Blackwell [20, 32].

Iatridou, Sabine (2000), 'The Grammatical Ingredients of Counterfactuality', *Linguistic Inquiry*, 31: 231–70 [16].

Jackson, Frank (1977), 'A Causal Theory of Counterfactuals', *Australasian Journal of Philosophy*, 55: 3–21 [8].

Khoo, Justin (2015), 'On Indicative and Subjunctive Conditionals', *Philosopher's Imprint*, 15/32 [15, 18].

Kratzer, Angelika (1986), 'Conditionals', in *Papers from the Parasession on Pragmatics and Grammatical Theory*, ed. A. M. Farley, P. Farley, and K. E. McCollough (Chicago: Chicago Linguistics Society) [3, 35].

Lappin, Shalom and Reinhart, Tanya (1988), 'Presuppositional Effects of Strong Determiners: A Processing Account', *Linguistics*, 26/6: 1021–38 [32].

Lewis, David (1973), *Counterfactuals*, Oxford: Blackwell [4, 7, 53, 66].

Lewis, David (1976), 'Probabilities of Conditionals and Conditional Probabilities', *Philosophical Review*, 85: 297–315 [36].

Lewis, David (1979), 'Counterfactual Dependence and Time's Arrow', *Noûs*, 13: 455–76 [67], Reprinted in Lewis 1986:

Lewis, David (1980), 'A Subjectivist's Guide to Objective Chance', in *Studies in Inductive Logic and Probability*, ii, ed. R. C. Jeffrey, Berkeley: University of California Press, 263–93 [75].

Lewis, David (1983), 'New Work for a Theory of Universals', *Australasian Journal of Philosophy*, 61: 343–77 [72], Reprinted in Lewis 1999:

McDermott, Michael (1996), 'On the Truth-conditions of Certain 'If'-sentences', *Philosophical Review*, 105: 1–37 [39].

Moss, Sarah (forthcoming), 'Subjunctive Credences and Semantic Humility', forthcoming in *Philosophy and Phenomenological Research* [75].

Nute, Donald (1980), *Topics in Conditional Logic*, Dordrecht: Reidel [45].

Pollock, John (1976*a*), *Subjunctive Reasoning*, Dordrecht: North-Holland [50].

Pollock, John (1976*b*), 'The 'Possible Worlds' Analysis of Counterfactuals', *Philosophical Studies*, 29 [50].

Rothschild, Daniel (forthcoming), 'A Note on Conditionals and Restrictors', in *Conditionals, Probability, and Paradox: Themes from the Philosophy of Dorothy Edgington*, ed. John Hawthorne and Lee Walters [35].

Stalnaker, Robert C. (1968), 'A Theory of Conditionals', in *Studies in Logical Theory: American Philosophical Quarterly Monograph Series, No. 2*, ed. Nicholas Rescher, Oxford: Blackwell, 98–112 [4], Reprinted in Harper, Stalnaker, and Pearce 1981:

Stalnaker, Robert C. (1976), 'Letter to van Fraasssen', in *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, i, ed. W. Harper and C Hooker, Dordrecht: Reidel, 302–6 [36].

Stalnaker, Robert C. (1978), 'Assertion', in *Syntax and Semantics 9*, ed. P. Cole, New York: New York Academic Press, 312–32 [10, 25], Reprinted in Stalnaker 1999:

Stalnaker, Robert C. (1981), 'A Defense of Conditional Excluded Middle', in *Ifs: Conditionals, Belief, Decision, Chance, and Time*, ed. William L. Harper, Robert Stalnaker, and Glenn Pearce, Dordrecht: D. Reidel Publishing Company, 87–104 [68].

Stalnakert, Robert C. (1975), 'Indicative Conditionals', *Philosophia*, 5: 269–86 [26, 28, 29].

Von Fintel, Kai (MS), 'Conditionals', in *Semantics: An international handbook of meaning*, ii, Handbücher zur Sprach- und Kommunikationswissenschaft 33.2, ed. Klaus von Heusinger, Claudia Maienborn, and Paul Portner, Berlin: de Gruyter Mouton, 1515–38 [36].

Von Fintel, Kai (1998), 'The Presuppositions of Subjunctive Conditionals', in *The Interpretive Tract*, MIT Working Papers in Linguistics, 25, ed. Uli Sauerland and Orin Percus, Cambridge, MA: MITWPL, 29–44 [10, 11].

Von Fintel, Kai (1999), 'NPI Licensing, Strawson Entailment, and Context Dependency', *Journal of Semantics*, 16: 97–148 [30, 64].

Von Fintel, Kai (2001), 'Counterfactuals in a Dynamic Context', in *Ken Hale: A Life in Language*, ed. Michael Kenstowicz, Cambridge: MIT Press [64].

Von Fintel, Kai (2004), 'Would you believe it? The King of France is back! Presuppositions and truth-value intuitions', in *Descriptions and Beyond*, ed. Marga Reimer and A. Bezuidenhout, Oxford: Oxford University Press [21].

Von Fintel, Kai (2007), 'If: The Biggest Little Word', Slides from a plenary address given at the Georgetown University Roundtable, `http://mit.edu/fintel/gurt-slides.pdf` [36].